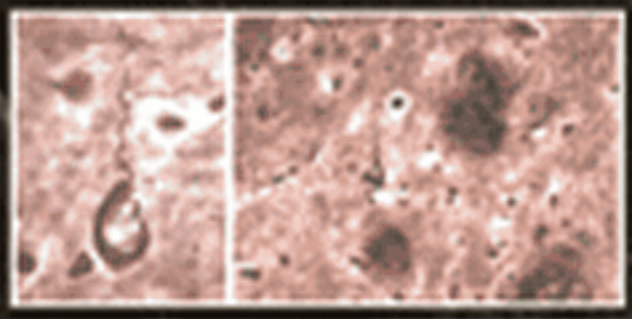
2006 U.S. Life Sciences Salary and Employment Survey >> Page 842

3 November 2006 | 312

Science



A Century of Research on

Alzheimer's Disease

AAAS

EDITED BY STELLA HURTLEY AND PHIL SZUROMI

Alzheimer's Disease After 100 Years

In Tübingen, Germany, on 3 November 1906, Alois Alzheimer described the first documented case of Alzheimer's disease, a neurodegenerative disorder that impairs memory, cognition, and behavior (see the cover). **Goedert and Spillantini** (p. 777) review what is known about the molecular pathology of the disease, which is defined by the presence within the brain of amyloid-β–rich plaques and tau-containing neurofibrillary tangles. **Roberson and Mucke** (p. 781) review the prospects for therapy to help delay or prevent pathological processes within the brains of afflicted individuals, in order to prolong patients' cognitive abilities, and maintain for as long as possible their quality of life.

A Need for a Sea Change

The significance of the ocean's declining diversity on humanity has been difficult to assess. In a series of meta-analyses, **Worm *et al.*** (p. 787; see the news story by **Stokstad**) quantify how the loss of marine diversity on local, regional, and global scales has affected the functioning and stability of marine ecosystems, the flow of ecosystem services, and the rise of associated risks to humanity. Similar relationships occur between biodiversity change and ecosystem services at scales ranging from small square-meter plots to entire ocean basins; this finding implies that small-scale experiments can be used to predict large-scale ocean change. At current rates of diversity loss, this analysis indicates that there will be no more viable fish or invertebrate species available to fisheries by 2050. However, the results also show that the trends in loss of species are still reversible.

Cosmic Shock Waves

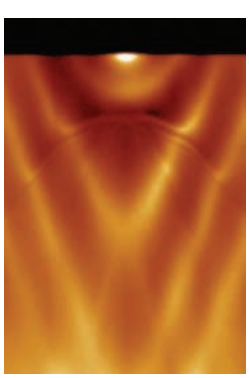
Clusters of galaxies grow by the infall of surrounding matter through gravitational effects, and peripheral shock waves are thought to be set up as material hits the cluster outskirts. **Bagchi *et al.*** (p. 791; see the Perspective by **Enßlin**) have used the Very Large Array to detect a ring of radio emission around a cluster that may signify such a shock wave. Giant twin radio arcs cup the cluster Abell 3376 and have size and brightness consistent with cosmological shock waves. The giant shock waves may provide sites for the acceleration of cosmic rays and particles associated with the structure-formation process.

Sublimation in Two Acts

Colloidal particles have been used as analogs for molecules for studying the formation of crystals or glassy jammed states. **Savage *et al.*** (p. 795; see the Perspective by **Frenkel**) use colloids to study the inverse problem, the sublimation of surface crystals. When the crystal is destabilized by a sudden temperature jump, large crystals first sublimate slowly. Once a critical size has been reached, the crystals suddenly melt into a metastable fluid before dissolving into the gas phase. In this regime of rapid change, the crystals were surrounded by a dense amorphous layer. The observations may correlate to the behavior of sublimating molecular systems and also to transitions in other systems like globular proteins.

Quantum Wells Run Deep

Thin metal films grown on semiconductor substrates can display quantum-well states created by electron confinement. Photoemission studies by **Speer *et al.*** (p. 804; see the Perspective by **Walldén**) of atomically uniform sliver films grown on Si(111) surfaces reveal additional electronic-fringe features for substrates with high levels of n-type doping. Despite the lattice mismatch and incommensurate growth of the Ag film, these features, which resemble diffraction waves, result from the electronic states from the film extending deep into the substrate and interfering with propagating states below the band edge. For lightly n-doped or p-doped samples, the band bending is too shallow to allow sufficient overlap of the film's states with those of Si.



From Heavy Fuels to Hydrogen

Biomass can be converted in oils and heavy liquids, but for many applications, it is desirable to further process these fuels into hydrogen or synthesis gas (a mixture of CO and H_2). However, the low volatility of these liquids often leads to long contact times with catalysts, and often some of the fuel is converted to carbon, which deactivates the catalyst. **Salge *et al.*** (p. 801) find that if heavy fuels such as soy oil or biodiesel are sprayed in the presence of O_2 as fine droplets onto already-hot rhodium-cerium catalysts, the heat of reaction causes the droplet to further break up and fully convert to H_2 (along with CO and CO_2, the other main products) without any added heat. These reactions are very fast (total contact times of 50 milliseconds or less), and no deactivation was seen after 20 hours of operation.

A Patchwork Solar Nebula

The use of the long-lived neodymium (Nd)–samarium (Sm) radioactive decay system for understanding very early processes in our solar system and Earth's differentiation depends on knowing the initial solar isotopic ratios. However, there is wide variation among measured Nd isotope levels among meteorites and compared with terrestrial samples. Two reports indicate that the early solar nebula was not well mixed with respect to this dating system or barium isotopes (see the 6 October news story by **Kerr**). Variations in Nd isotopes seen between chondrites and earth samples led to the suggestion that some Nd isotopes were sequestered deep in the earth. **Andreasen and Sharma** (p. 806) have measured Nd and Sm isotopes in primitive carbonaceous chondrite meteorites and find that

Continued on page 723

CREDITS (TOP TO BOTTOM): UNIVERSITY CLINIC OF TÜBINGEN; SPEER *ET AL.*

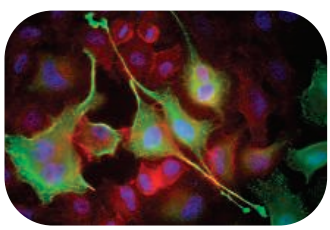
Published by AAAS

Continued from page 721

the variations among meteorites are real and are caused primarily by a p-process (photodissociation of nuclides) deficit in carbonaceous chondrites relative to ordinary chondrites, eucrites (from Vesta), and the terrestrial standard. **Ranen and Jacobsen** (p. 809) have measured barium isotopes in chondrites and found that they also exhibit variations among meteorite types, which they interpret as implying that the protosolar nebular was heterogeneous. Chondritic meteorites originated in a different place and were more enriched in supernova-derived material compared with Earth.

Pores, Sieve, or Gel?

Nuclear pore complexes behave like a sieve—they allow small molecules to pass freely but restrict passage of macromolecules (>30 kilodaltons) between the nucleus and cytosol. The so-called FG-rich nucleoporin repeats, which are intrinsically unfolded protein domains that contain short clusters of hydrophobic amino acids separated by hydrophilic spacer regions, are thought to form the barrier, but the functional organization of this barrier has remained a matter of speculation. **Frey *et al.*** (p. 815; see the Perspective by **Burke** and the Perspective by **Elbaum**) show that these FG-rich repeats occur in an extended conformation and form a noncovalent (and thus reversible) hydrogel. Hydrophobic bridges connecting the individual polypeptide chains and create a three-dimensional sieve-like structure that is crucial for nuclear pore complex function.



Neurite Extension and Membrane Trafficking

Neurons extend processes (neurites) that can reach more than 1 meter in length. Formation of neurites requires both cytoskeletal remodeling and membrane transport. **Shirane and Nakayama** (p. 818) have now identified a protein, protrudin, that is essential for neurite extension triggered by nerve growth factor. Protrudin binds to and inhibits the activity of the protein Rab11, which functions as a molecular switch in membrane transport to the cell surface.

A Cool Way to a Long Life

Caloric dietary restriction prolongs life span in a variety of organisms, and in mammals the resultant lowering of core body temperature has been offered as one potential explanation. **Conti *et al.*** (p. 825; see the Perspective by **Saper**) generated transgenic mice that overexpress mitochondrial uncoupling protein 2 in hypocretin-producing neurons within the hypothalamus, which lowers core body temperature by about 0.5°C. In the absence of caloric restriction, the median life span of these "cool mice" was about 15% greater than that of their wild-type littermates.

Life Isn't Fair

In the two-player ultimatum game, low offers made by the first player to divide the pot of money unequally are considered unfair. The second player can choose either to accept these low offers (in which case the first player walks away with more than half of the pot) or to reject them (in which case neither player receives anything), with the outcome reflecting the competition between selfishness and indignation. The dorsolateral prefrontal cortex (DLPFC) is thought to play a role in the decision-making process. **Knoch *et al.*** (p. 829, published online 5 October) used repetitive transcranial stimulation to test this directly by interfering with DLPFC function. Suppressing DLPFC activity tilted the competitive motivations toward the side of selfish behavior and a greater acceptance of unfair offers.

Donning the Myelin Sheath

The myelin sheath electrically insulates axons and makes the conductance of neuronal impulses much more efficient. **Chan *et al.*** (p. 832) examined how the Schwann cells begin myelination of an axon. In neuronal cultures, a cell polarity protein, Par-3, localized to where the Schwann cell meets the axon and promoted the recruitment of the brain-derived neurotrophic factor (BDNF)–receptor to the junction between Schwann cell and neuron. BDNF-dependent signaling between the axon and Schwann cell then ensures that myelination begins. This localized signaling arising from Par-3 polarity may help to ensure that the myelination begins at the right spot.

CREDIT: SHIRANE AND NAKAYAMA

Published by AAAS

Donald Kennedy is the Editor-in-Chief of *Science*.

Salary Survey

EVERY FEW YEARS SINCE 2001, *SCIENCE* HAS CONDUCTED A SURVEY IN WHICH LIFE scientists in the United States report how well they're doing financially—and better yet, how they are feeling about their profession and their place in it. This year's report (p. 842) contains some elements of relief for those who have chosen to do some kind of biology for a living. The relief could hardly come at a better time. Earlier signals have included lowering pay lines for National Institutes of Health (NIH) grants, unionization campaigns among postdoctoral fellows, and the grim prospects for "domestic discretionary" expenditures (which include, of course, research funding). It's tough when the first NIH grant in one's career comes after the age of 40.

But the survey shows that life scientists at all levels are doing better than in recent years and better than inflation. Full-time academic life scientists earned 5.4% more this year than in the preceding year, well above the cost-of-living index. Postdoctoral fellows, who once were used to feeling like a disfavored class, did even better, with an average salary increase of 8.1%. Ph.D.s who work in industry continue to earn more than their academic counterparts, by an average of about $40,000 (their salaries are increasing faster, too, up 10% this year). It may not surprise readers that job satisfaction shows a positive relationship with compensation. But the linkage is weaker than one might expect. The job satisfaction of the top group of earners is high all right, but the group earning only one-fifth as much reports only slightly lower satisfaction. It appears that prestige, promotion opportunity, and intellectual challenge are more important determinants. No surprise there.

Underneath all this good news, however, lie some significant submerged inequities. For example, the average salary for academic pharmacologists is about $55,000 more than for developmental biologists. There is a warning signal that accompanies disparities of that kind: Scientists, like most other kinds of workers, compare salaries, and when they are disappointed in the results, morale is likely to decline and complaints are certain to follow. That can spell trouble in an institution. One example is the often-remarked academic salary differential between professors of law and English, which led one of the former to deliver this unsympathetic advice to a plaintive colleague in comparative literature: "Well then, go out and practice English."

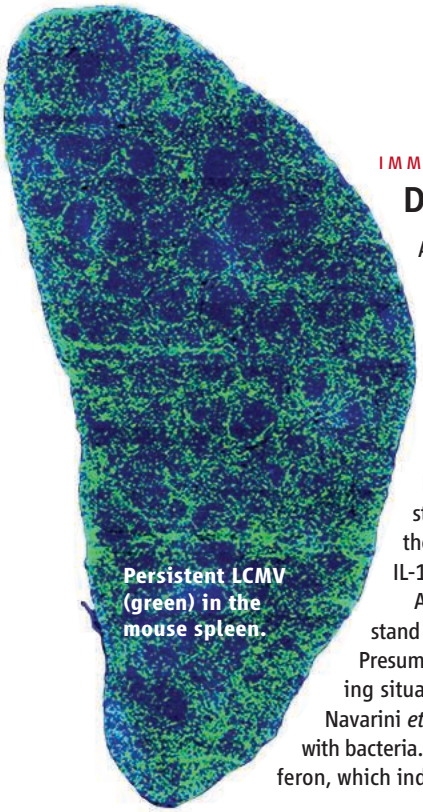
It is also important to note that the improved status of our sample of scientists actually serves to widen an already growing gap in our national economy. We should be worrying about the rewards and satisfactions of our scientific colleagues, but we should also be concerned about the people who clean their labs, run the cafeteria, and work in the accounting office. That brings us to some discomfiting facts about the pattern of wage changes in the contemporary economy. Improvement, often faster than inflation, is seen in the upper range of the wage scale, especially for employees in the service economy. But at the lower end, workers are worse off. This seems paradoxical, because productivity is up: If these workers are doing more, why aren't their wages keeping pace?

One answer is that, increasingly, full-time workers are being replaced by part-timers. Large employers like Wal-Mart are elevating the proportion of their employees who are part-time, and there and elsewhere, outsourcing strategies are getting more and more popular. The result is that those who remain are having to take less, while their outsourced replacements make a little more but lose their benefits. Organized labor once provided a countervailing force, but unions have lost much of their strength, and management is taking advantage of the weakened opposition.

None of this should take away from our good feeling about the improved prospects for our colleagues in the life sciences. This decade, after all, has seen remarkable progress in biomedical research, and that is surely a very important enterprise for which its practitioners deserve a fair reward. But all of us, scientists included, will benefit from a society that enjoys a stable political economy, and a more equitable income distribution can contribute to making that a more sustainable benefit.

–Donald Kennedy

10.1126/science.1136531



Persistent LCMV (green) in the mouse spleen.

IMMUNOLOGY

Double Jeopardy

A persistence of infection is a trademark of some viruses, notably HIV and hepatitis B and C viruses, and is in large part the result of an incompletely effective antiviral T cell immune response. Brooks *et al.* provide evidence that a familiar and important regulatory cytokine—interleukin (IL)–10—plays a key role in facilitating viral persistence. Using two strains of the mouse lymphocytic choriomeningitis virus (LCMV), one capable of establishing a persistent infection and the other not, they showed that the levels of IL-10 generated during infection with the former were significantly higher than with the rapidly cleared strain. The T cell responses in the persistently infected animals were also diminished, leading them to ask if IL-10 were directly responsible for allowing the persistence of one strain (and not the other) by dampening virus-specific T cells. Indeed, in mice lacking IL-10, the strain differences were less apparent and, as they also found using blocking antibodies for IL-10, this led to faster viral clearance and signs of improved T cell memory.

Although differential IL-10 expression might explain such effects, it will be useful to understand exactly how the two different strains of virus trigger distinct levels of IL-10 in the first place. Presumably, mechanisms that diminish immunity could influence other ongoing infections, including situations of bacterial co-infection. Although IL-10 was not itself tested in another study by Navarini *et al.*, these authors do report that LCMV increases the susceptibility of mice to co-infection with bacteria. Rather surprisingly, the culprit in this case was the innate antiviral cytokine type I interferon, which induced apoptosis in bacteria-clearing granulocytes. — SJS

Nat. Med. **12**, 10.1038/nm1492 (2006); *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15535 (2006).

MATERIALS SCIENCE

Turning Water Inside Out

Colloidal particles with appropriate surface properties adsorb strongly at liquid/liquid and vapor/liquid interfaces; hence, they are used as stabilizers for emulsions and foams. Particle surface wettability can be tuned to entrap water in oil, or oil in water, for example, and even to switch between these two regimes. In the case of a vapor/liquid interface, such inversion behavior—the shift from air bubbles dispersed in water, as in a foam, to water droplets dispersed in air—has been explored only recently. Binks and Murakami stabilize a full range of air/water dispersions by adding silica particles 20 to 30 nm in diameter that vary in their wettability, which the authors reduce by lowering the concentration of surface silanol (SiOH) groups via hydrophobic capping. High SiOH content gives rise to stable aqueous dispersions, whereas intermediate particle hydrophobicity leads to air-in-water foams. At the lowest SiOH content, the particles drive a transitional inversion, coating discrete water droplets to stabilize a water-in-air powder. This powder releases water to the skin when sheared by rubbing, suggesting possible applications in cosmetics. The authors further show that varying the ratio of water to air at fixed SiOH content can also force an inversion (in this case formally termed "catastrophic"), giving rise to a soufflé-like material stickier and more highly aggregated than the water-in-air powder. — MSL

Nat. Mater. **5**, 10.1038/nmat1757 (2006).

MICROBIOLOGY

Intracellular Demographics

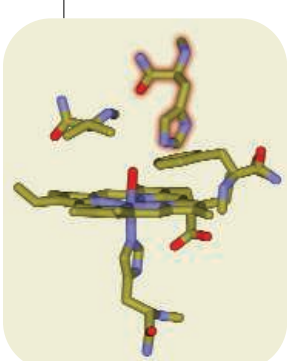
Mathematical models are commonly used to help predict the course of epidemics through a population of organisms. In contrast, Brown *et al.* have absorbed recent findings on intracellular events in salmonellosis to develop a within-organism model. *Salmonella enterica* grows within host phagocytes to varying cell density regardless of cell "permissiveness." Interestingly, the model hints that apoptosis of host cells has little effect on controlling the spread of infection; rather, the driving force is necrotic bursts that release bacteria to spread into new foci of infection. Comparing attenuated (i.e., vaccine) and virulent strains of salmonellae in the model reveals that an attenuated strain replicates less well than a virulent strain and uses the same number of host cells to do so; hence, the resulting pathology may be rather similar. The model is clearly useful for predicting the effect of combination drug therapy, and it hints that drugs that kill extracellular pathogens might select for "refuge resistance"; that is, the suppression of cell lysis mechanisms. — CA

PLoS Biol. **4**, e349 (2006).

CHEMISTRY

Embedding a Reporter

The nitrile (CN) group can be a useful infrared reporter in proteins because it has a strong stretching vibration near ~2200 cm^{-1}, a spectral region usually free from interfering absorptions in a biochemical environment. Schultz *et al*. have devised a protocol to introduce the non-naturally occurring amino acid *para*-cyano-L-phenylalanine (pCNPhe) into proteins during bacterial synthesis, using an orthogonal nonsense suppressor transfer RNA (tRNA) paired with an aminoacyl-tRNA synthetase derived from *Methanococcus jannaschi*. They apply this system to incorporate the pCNPhe reporter in place of a histidine residue (His^{64}) near the ligand-binding site of the heme group in myoglobin. When water was bound in the active site, they observed an 11-cm^{-1} shift in frequency relative to pCNPhe absorption in pure water or buffer solution, a



His^{64} (highlighted) in myoglobin was replaced by pCNPhe.

Continued on page 729

CREDITS (TOP TO BOTTOM): BROOKS *ET AL.*, *NAT. MED.* **12**, 10.1038/NM1492 (2006); SCHULTZ *ET AL.*, *J. AM. CHEM. SOC.* **128**, 10.1021/JA0636690 (2006)

Continued from page 727

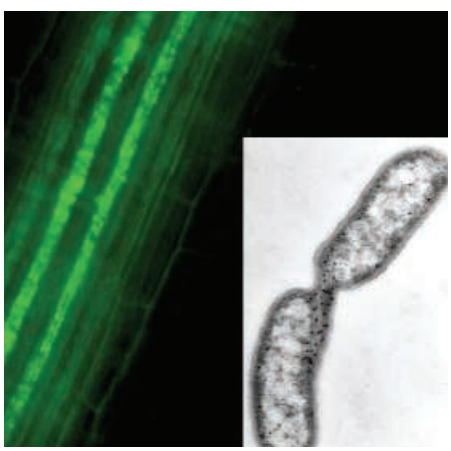
change consistent with increased water polarity in the binding pocket. Changes in the observed CN stretching frequency were also consistent with the bent conformations of Fe(II)-bound NO and O_2, as well as the linear CO-Fe(II) complex. — PDS

J. Am. Chem. Soc. **128**, 10.1021/ja0636690 (2006).

BIOTECHNOLOGY

Keeping the Fix In

The endophyte *Azoarcus* sp. strain BH72 resides within the roots of rice and other grasses. In return for supplying the plant with fixed nitrogen (diatomic nitrogen that has been converted into biochemically tractable forms such as ammonia), it is presumed to benefit from a sheltered and predictable habitat. Krause *et al.* have sequenced its genome and compared it to that of a free-living relative, the strain EbN1. They find the expected suite of nitrogen-fixing and -metabolizing enzymes along with a large set of transporters for dicarboxylic acids (though not of sugars) and chelated iron. On the other hand, the low-stress lifestyle appears to have led to the loss (or non-acquisition) of type III and IV secretion systems as well as a paucity of virulence and pathogenic components. Similarly, there are only a small number of mobile elements, in comparison to its independent cousin. How these characteristics might be harnessed in agronomic efforts to enhance rice cultivation, and perhaps that of other cereals, is not yet clear, but it's a start. — GJC



***Azoarcus* in a rice root.**

CREDIT: THOMAS HUREK

Nat. Biotechnol. **24**, 10.1038/nbt1243 (2006).

PSYCHOLOGY

Theorizing Takes Time

The human ability (commonly referred to as a theory of mind) to formulate inferences about the mental states, such as beliefs and intentions, of others is a fundamental item in our social cognitive skill set. Apperly *et al.* have asked whether the component reasoning processes operate in an automatic fashion, in the background as it were, and yield output that can be summoned effortlessly when needed. Using the canonical Maxi type of false-belief task (which some might argue has attained a mythic status), they required that participants report the final positions of objects, both their actual locations as well as where female actors believed them to be. In comparison to keeping track of the physical objects, participants required more time to infer where the actors thought they were located; though if explicitly forewarned to monitor belief states, they were equally fast at specifying actual and supposed locations. Does this mean that we do not automatically maintain a running tally of who believes what? Not quite—there may be an unconscious pre-processing of evidence into candidate belief states, where the final step of asserting which one to act on is taken only on demand. — GJC

Psychol. Sci. **17**, 841 (2006).

APPLIED PHYSICS

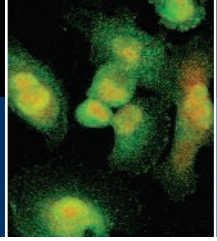
Scaling Superconductive Memories

Superconducting electronic systems offer great potential to improve the speed of conventional computers through low power dissipation and switching times on the order of picoseconds. One problem, however, has been to develop small-sized memory storage elements that are also compatible with large-scale integration. For instance, data storage in these systems has generally been based on harboring magnetic flux in a superconducting loop, and those loops tend to be several micrometers in diameter.

Held *et al.* propose the design of a memory element based on a ferromagnetic dot coupled to a superconducting Josephson junction. Because the critical current of a Josephson junction is magnetic field–dependent, the magnetization of the dot can be switched to modulate the field in the junction either below or above a critical value. The data, 0 or 1, are thus stored as the magnetization direction in the dot and can be read out as the critical current of the Josephson junction. Preliminary experiments using a Permalloy ($Ni_{81}Fe_{19}$) dot demonstrate the principle of operation and also show nonvolatile storage capability at room temperature. The authors note that optimization of the device should reduce the relatively high (~100 mA) applied currents required to switch the dot magnetization. — ISO

Appl. Phys. Lett. **89**, 163509 (2006).

Published by AAAS



A surprising role for RNA 741

New H5N1 strain spreads 742

NASA SPACE SCIENCE

Hubble Gets a Green Light, With Other Missions on Hold

Starry-eyed. NASA's Mike Griffin tells Goddard scientists about plans to repair the Hubble Space Telescope.

Relations between the science community and NASA chief Mike Griffin are at best frosty. But this week, he won enthusiastic applause from delighted astronomers at Goddard Space Flight Center in Greenbelt, Maryland, by promising to send astronauts back to the aging Hubble Space Telescope in May 2008 to extend its operating life well into the next decade.

The announcement ended nearly 3 years of rancorous debate among politicians, scientists, and engineers over whether the orbiting satellite should live or die. "It's been a long time coming. … It's a great day for science," said Senator Barbara Mikulski (D–MD), who led the fight to save Hubble. But the $350 million servicing mission will make it even harder for NASA to fund future astronomy missions.

Shuttle astronauts have visited Hubble four times since its launch in 1990, each time swapping instruments, replacing batteries, and performing other maintenance tasks. Those challenging space walks—including the first mission in 1993 to fix Hubble's faulty main mirror—also helped to improve the quality of data beamed back to Earth. A fifth and final servicing flight was planned for 2004, although scientists were pressing for a sixth mission later in the decade when a returning Columbia disintegrated over Texas on 1 February 2003.

The following year, then-NASA chief Sean O'Keefe canceled the fifth servicing mission because of safety concerns. Hubble circles Earth in a different orbit from the space station. If the astronauts were to face an emergency during a Hubble visit, the crew would not be able to reach the space station and wait for rescue by another orbiter. O'Keefe argued that the possible loss of lives was not worth the additional scientific results from Hubble. But a chorus of scientists and politicians—in particular Mikulski—raised a ruckus.

Seeking a compromise, O'Keefe proposed a robotic repair mission. But a National Academy of Sciences' panel rejected that idea as technically too difficult, costly, and time-consuming. It also urged NASA to reinstate the shuttle mission, recommending that the science program not bear all of the expected $1 billion cost of the mission.

Taking over from O'Keefe in April 2005, Griffin pledged to reverse his predecessor's decision if subsequent shuttle flights demonstrated that the fleet could be operated safely. "What's different now is that we have three flights under our belt," says Goddard Director Ed Weiler. Those successful flights, Griffin told scientists this week, have allowed him to reverse a "troubling, troubled, and unpopular decision."

Griffin's decision means that NASA will spend most of its astronomy budget on three major missions—the Hubble servicing flight, construction of the James Webb Space Telescope, and the Stratospheric Observatory for Infrared Astronomy (SOFIA). Technical troubles, schedule delays, and cost overruns plague the latter two. But Weiler says that the Webb is back on track after a rough couple of years, while SOFIA—which Griffin initially canceled only to revive in July—is slated to begin operations in 2009. Those large projects leave little room for smaller or future missions. For example, NASA halted work earlier this year on the extrasolar planet-seeking Space Interferometry Mission (SIM) in order to cover SOFIA's cost overruns. Those pressures worry some astronomers, who fear that the three missions will limit new efforts.

"Is the astronomy program with just [Webb], Hubble, and SOFIA a good astronomy program? You betcha," says Weiler. Although he acknowledges that there is a gap in smaller missions for the next few years, he notes that the cost of building the Webb will peak in 2008 and then decline over the next 5 years. "The big issue now is what to do with that wedge."

The four leading contenders appear to be the Joint Dark Energy Mission with the Energy Department, a mission called Constellation-X that features a bevy of x-ray telescopes, the Laser Interferometer Space Antenna to study black holes and the early universe, and SIM. NASA had intended to fund all in this decade and the next, but budget constraints likely will make for a competitive race.

Weiler also urged scientists to think about smaller, less costly missions. He is pressing to build smaller satellites that could be launched from the Wallops Flight Facility on the coast of Virginia, using converted U.S. military missiles. The 11 December launch of an Air Force satellite could mark the start of a series of smaller missions. In the meantime, "astronomers are elated at the NASA decision," said Steve Maran, spokesperson for the American Astronomical Society. "It's fantastic," adds Mario Livio, an astronomer at the Space Telescope Science Institute in Baltimore, Maryland. "Clearly, we are ecstatic."

–ANDREW LAWLER

CREDIT: PAT IZZO/NASA GODDARD FLIGHT CENTER

U.S. POLITICS

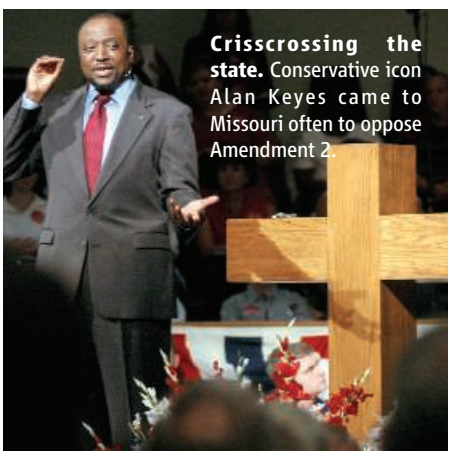
Scientists Look to Missouri to Show the Way on Stem Cells

ST. LOUIS, MISSOURI—"I'm leaping into deep waters this morning," Reverend James Morris confessed as he began his sermon. His topic, "Stem Cells: What Would Jesus Say?" isn't something that the pastor of Lane Tabernacle CME Church, a white-stoned building in a beat-up, African-American neighborhood in the north side of town, had ever discussed before from the pulpit. But today—Sunday, 29 October—was different. His congregants, and their fellow citizens across this Midwestern state, had been bombarded for weeks with television advertisements for and against Amendment 2, a ballot initiative instigated by Missouri scientists to allow human embryonic stem cell research in the state while banning reproductive cloning. With barely a week left before the election, and with both sides claiming the moral high ground, Morris decided it was time to advise his flock.

The barrel-chested pastor moved easily from the sacred to the profane. "I come that they might have life," Morris cried out, quoting the gospel of John to amens and hallelujahs from the pews. Then, in more measured tones, he defined stem cells and explained that embryos obtained by fertility treatments are generally discarded. "I say, use them for cures for diabetes, use them for cures for sickle cells, cures for heart disease." And later, as the pianist segued into a gospel tune: "Vote your conscience on November 7, after prayer and reflection."

Reverend Morris's words are part of a long-running debate in the Show-Me state. In 2004, Missouri's small science community, anchored in Kansas City and St. Louis, has sought state funding for the expansion of private biomedical research. But each year legislators, backed by the states' potent prolife lobby, scuttled the bills by threatening to include a ban on all forms of human cloning, including a technique, somatic cell nuclear transfer, that is useful for producing stem cells.

Last year researchers, the biotech community, and patients' groups proposed taking the issue directly to voters in the form of an amendment to the state constitution. Amendment 2 would make cloning "for the purpose of initiating a pregnancy" a felony, and it would bar state legislators from prohibiting other kinds of stem cell research with embryos, including somatic cell nuclear transfer (SCNT). Supporters have raised an unprecedented $30 million to promote the amendment. The biggest donors are James and Virginia Stowers of Kansas City, philanthropists and cancer patients who have poured the bulk of their wealth into the $1.6-billion Stowers Institute for Medical Research in that city.



Crisscrossing the state. Conservative icon Alan Keyes came to Missouri often to oppose Amendment 2.

On the spot. Actor Michael J. Fox attacked incumbent Republican Senator Jim Talent's opposition to Amendment 2 in ads for Democrat challenger Claire McCaskill.

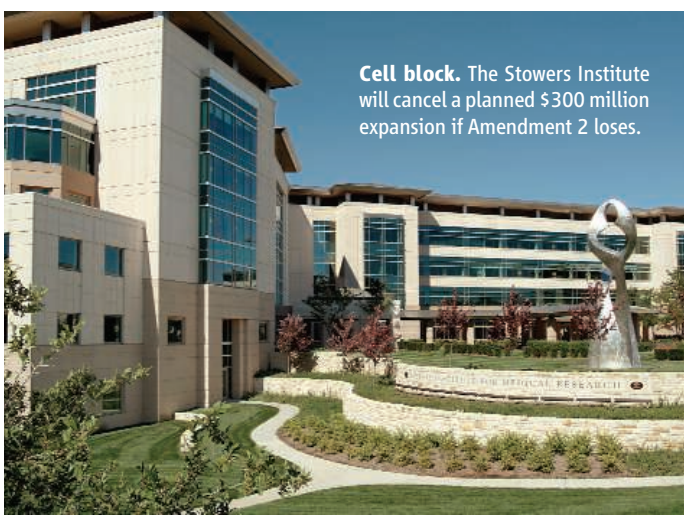
Supporters of the amendment, which include business and disease advocates, have spent most of their war chest on television, radio, and billboard advertisements touting the potential for cures. But there are also footsoldiers, including a few dozen scientists and doctors. Washington University (WU) bone pathologist Steven Teitelbaum estimates that he's delivered a power-point talk once a week for 2 years to community and church groups across the state. His colleague, soft-spoken James Huettner, who is one of the few Missouri scientists using human embryonic cells, has described his work with neurons at several area churches. The outreach role is "a new thing," says Huettner, but necessary for what Huettner and others hope may eventually lead to new treatments for Alzheimers' patients.

If Amendment 2 fails, Missouri scientists fear a worsening in what is already a challenging environment in the state. "It would really be disappointing if this thing doesn't pass and we can't do [embryonic] stem cell research," says Washington University graduate student Katherine Varley, who says she declined offers from molecular biology labs on both coasts "because the [WU] faculty is so good." Managers at Stowers want desperately to fund work in human embryonic cells and have pledged to cancel a planned $300 million expansion in Missouri if the amendment fails. The philanthropists have also given Harvard scientists $11 million for work that they would have preferred to see done at their own institute. The same year, Illinois Governor Rod Blagojevich invited 30 Missouri doctors and researchers to cross the Mississippi River and continue their stem cell-related work in his state.

In September, after an ad barrage featuring clergy, doctors, firefighters, and singer Sheryl Crow extolling its virtues, the amendment had a 21% lead in a *St. Louis Post-Dispatch* poll, and only 5% of voters were undecided. But the opposition is making strides: This week, the paper reported the lead had slipped to 17%—and that 14% of voters hadn't made up their minds. Teitelbaum says researchers "are not taking anything for granted," and are counting on local leaders such as Reverend Morris to shore up support. ▶

CREDITS (TOP TO BOTTOM): KELLEY MCCALL/AP PHOTO; MCCASKILL CAMPAIGN/AP PHOTO

Outspent by a margin of 10 to one, the largely religious and conservative groups that make up the opposition are relying on their grassroots supporters to carry their message. The Bott Radio Network's 17 evangelical radio stations in Missouri have been playing ads opposing the amendment—including Catholic, Baptist, and Lutheran voices—three times an hour for months. The opposition's message: Amendment 2 would sanction both the destruction of embryos and the creation of life. "Both of them are intrinsically evil," thundered well-known evangelical minister Rick Scarborough of Texas' Vision America at a rally of 150 in St. Louis last week, the last of six events around the state to recruit Missouri's diverse clergy to the cause. A number of prolife researchers have given scientific heft to their arguments, including Richard Chole, a WU bone researcher, who has kept a vigorous schedule of evening talks and debates statewide. "I never get home anymore," he says.



Cell block. The Stowers Institute will cancel a planned $300 million expansion if Amendment 2 loses.

CREDIT: DON IPOCK PHOTOGRAPHY

Other opponents have made more secular arguments. Cathy Ruse, who works with the Family Research Council in Washington, D.C., and was brought in to assist Missourians Against Human Cloning (MAHC), told *Science* that Scarborough's event was "not our rally." Instead of emphasizing dogma, Ruse and others have argued recently that Amendment 2 is simply deceptive. The law would make it illegal for Missouri lawmakers to bar somatic cell nuclear transfer. In a procedure sometimes called therapeutic cloning, researchers would like to use SCNT to transplant DNA into embryos from which they could derive stem cells genetically matched to patients. But SCNT was also the first step Ian Wilmut and his colleagues used to create Dolly, which leads Chole and other opponents to brand proponents "disingenuous" when they tout the amendment as a cloning ban.

MAHC has also attacked amendment language that would legalize "reimbursement for reasonable costs" for egg donors, calling it exploitative of women. Harvard ethicist Louis Guenin says the amendment's language in this area follows ethical guidelines laid out by the American Society of Reproductive Medicine; Teitelbaum says the debate over the amendment's narrow definition of illegal cloning is "semantic" and that his side makes clear what it would bar and allow. The initiative may also affect a dead-even race for Missouri's U.S. Senate seat.

The challenger, Democrat Claire McCaskill, has emphasized her support for embryonic stem cell work with an advertisement featuring actor Michael J. Fox, visibly tremulous from Parkinson's disease. The ad, which made national news, called on the incumbent, Republican Jim Talent, to drop his opposition and help "millions of Americans. American's like me." Fox and others hope a win in Missouri could pave the way for federal support for work with embryos—research that states have taken up since President George W. Bush's 2001 announcement barring federal funding for new embryonic cell lines. A recent poll suggests half of Talent's supporters favor Amendment 2.

Missourians are used to tight contests in this oftentimes swing state, and the Amendment 2 race could also come down to the wire. "I think it'll pass, but it will be close," Dorothy Cartwright told a fellow congregant as they examined pro-amendment flyers on their way out of church.

Both sides say a lot is at stake. "If [Amendment 2] does not pass, [it's] likely Missouri will become the Kansas Board of Education, part two, for the nation," says prominent St. Louis attorney Walter Metcalfe, referring to that body's repeated attempts to remove evolution from the school curriculum. But Teitelbaum says that not putting the issue on the ballot would have been a worse strategy. "It's a risk that had to be taken," he says.

–ELI KINTISCH

All Bent Out of Shape at *Topology*

In the latest brouhaha over journal prices, all nine members of the editorial board of *Topology*, a prestigious math journal based at the University of Oxford, U.K., plan to step down at the end of the year to protest the rising cost of institutional subscriptions.

Published six times a year, the title costs €100 in Europe for individuals and €1488 for institutions. In a 10 August letter to publisher Elsevier that has recently drawn media attention, the editors said the prices have had a "damaging effect" on *Topology*'s reputation. Elsevier says it has "moderated" price hikes, but since 1999, editors of at least two other Elsevier journals have stepped down in a similar protest. **–JOCELYN KAISER**

Investigating the Investigators

CAMBRIDGE, U.K.—Concerned about the expanding use of human DNA in criminal investigations, the U.K. Nuffield Council on Bioethics announced this week that it will examine how this forensic tool might endanger privacy and fair legal procedures. The review will feature public comments and a panel with legal and scientific experts.

Britain is ripe for the one-year review, says chair Bob Hepple, emeritus professor of law at the University of Cambridge, because it maintains a "virtually unregulated" forensic DNA database—the world's largest. The bank holds 3.45 million entries taken from suspects, crime scenes, victims, witnesses, and volunteers who wanted to be excluded from inquiries. Hepple says it's not clear how a citizen may remove DNA from the bank, which now covers 5% of the U.K. population. "[T]he issue is whether our DNA belongs to us or to the state," says Hepple. **–ELIOT MARSHALL**

Back to School

The number of foreign students enrolling in U.S. graduate schools this fall has jumped by 12%, according to a new survey by the Council of Graduate Schools (CGS). The rise indicates an accelerating recovery for international graduate enrollments, which posted a 1% increase last year after declining for 3 years. Enrollments from India this year leapt 32%; enrollments from China jumped 20%. U.S. academic officials, attribute the increase to streamlined visa procedures by the government and increased outreach by U.S. institutions. "This encouraging trend will continue," predicts CGS president Debra Stewart.

–YUDHIJIT BHATTACHARJEE

RESEARCH FUNDING

Olympics-Level Costs Upset Plan to Move U.K. Biomedical Institute

A plan to relocate Britain's largest biomedical research unit is running into a big and—some would say—predictable problem: London's high prices. The National Institute for Medical Research (NIMR), known for its work on infectious diseases, is slated to move from its 1950s suburban digs to a modern building in the center of London by 2012. But its parent agency, the Medical Research Council (MRC), disclosed last week that this plan is being reworked to reduce its cost. The price of the in-town building has gone up so much—to about $698 million, roughly $89 million more than projected—that U.K. treasury officials have balked, according to some NIMR staffers.

When first proposed, the idea of moving NIMR from Mill Hill to the city drew criticism from some of NIMR's 700-plus staff. Several well-known researchers said they were concerned that facilities and personnel might be shed to make the move affordable (*Science*, 4 February 2005, p. 652). Staying put at Mill Hill might be preferable, they argued.

Scalpel, please. MRC chief Colin Blakemore has been asked to wield the budget knife.

The MRC, led by chief executive Colin Blakemore, responded that the move was essential because the agency's basic researchers needed to build closer ties with clinicians. In 2005, the MRC forged a partnership with University College London and its hospital and bought land nearby. Blakemore said a move would not significantly reduce NIMR's research corps or facilities.

Last week, however, the journal *Research Fortnight* reported that the U.K. treasury, alarmed about the rising cost of the NIMR project, was refusing to release funds needed to finance construction.

This prompted a sharp rebuttal from Blakemore. "The Treasury has *not* rejected a bid for additional funding for the proposed move of NIMR into central London," he wrote on 25 October. In fact, he said, the case for construction has not been formally submitted for approval. Blakemore confirmed, however, that the price of the new NIMR has gone up, mainly because of "revised projections for building work in London." (NIMR will be competing for skilled workers with the 2012 London Olympics.) NIMR's new estimated cost of $698 million "exceeds all funds that are potentially available," Blakemore reported. The MRC council concluded last month that expenditures must not exceed the NIMR's current budget of $65 million per year. It also concluded that the high London costs and the higher fraction of the budget going to clinical expenses would require a reduction in staff. At the same time, it ruled out staying at Mill Hill. Blakemore said staff cuts would be achieved through normal "turnover in the coming 5 years."

Iain Robinson, head of molecular neuroendocrinology at NIMR who has been involved in discussions of NIMR's future, says, "We will have to make a better case" for investing in the move to London. "It's an unhappy situation to be in, but it's good to have the government's position clarified," Robinson says. The MRC Council is expecting to review a revised NIMR plan on 13 December and present it for government approval in late January.

–ELIOT MARSHALL

CONFLICT OF INTEREST

NIH Rules Rile Scientists, Survey Finds

A staff survey at the National Institutes of Health (NIH) reveals that intramural scientists have strong negative feelings about the agency's strict new ethics rules. But whether those rules are triggering a flight from the NIH campus in Bethesda, Maryland, is harder to measure.

The Web-based survey,* conducted this summer by an outside contractor, examined employee morale since NIH imposed new ethics rules in August 2005. Those rules followed a series of newspaper stories and congressional hearings about senior scientists who had received large consulting fees from companies. The new rules bar NIH employees from undertaking paid consulting for industry, restrict ownership of drug company stock, and limit awards.

Those limits go too far, say most of the 512 tenured and tenure-track researchers who filled out the survey. (That's out of a pool of roughly 1200 and doesn't include support scientists.) A majority (57%) agree the ethics rules needed to be addressed, 80% of respondents now find them too restrictive, and roughly 90% worry that they will harm NIH's ability to recruit and retain staff.

But the message becomes murkier when the survey hits closer to home. Some 39% of tenured and tenure-track staff say the new rules are leading them to look for or consider finding work outside NIH. At the same time, 79% say they are happy with their jobs, and 86% say they expect to be at NIH next year.

NIH Deputy Director Raynard Kington says any change will cause some employees to "think about" leaving, but that doesn't mean they will. In a staff memo last week describing the survey results, NIH Director Elias Zerhouni noted that although "the survey does suggest concerns" about recruitment and retention, attrition rates for all scientific staff have remained steady for the past few years.

NIH intends to further analyze attitudes to the new rules among scientists who left NIH recently and those considering a move to the agency. "When we feel there is a strong case, we'll be the first to advocate changing [the rules]," Kington says. "We're not at that point yet."

NIH bioethicist Ezekiel Emanuel, who is a member of the NIH Assembly of Scientists, argues that the 39% who are considering leaving NIH is "an enormously high rate." He claims that the rules have hindered recruitment, pointing to several senior positions at the National Cancer Institute that have gone unfilled for a couple of years. "We need a more rational policy and a less cumbersome policy," Emanuel says.

–JOCELYN KAISER

* http://www.nih.gov/about/ethics/evaluationslides.pdf

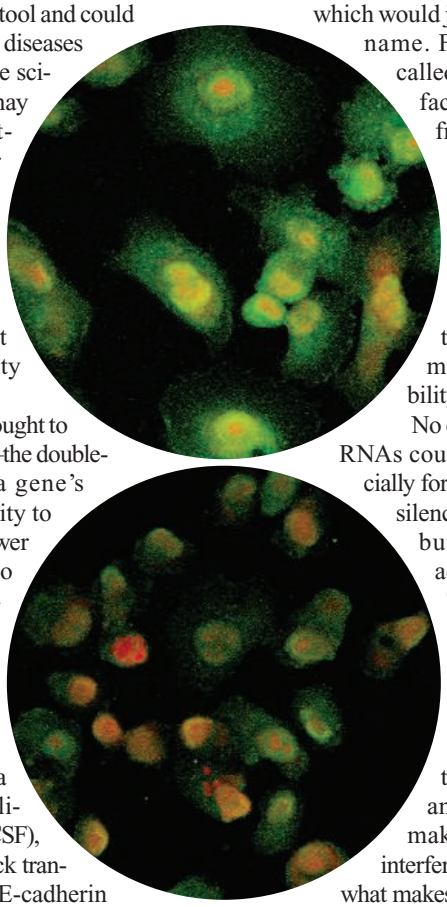
CREDIT: MRC

GENETICS

Small RNAs Reveal an Activating Side

The ability of short double strands of RNA to turn off specific genes, a process called RNA interference (RNAi), has enabled new animal models, spawned biotech companies, and a few weeks ago, produced a Nobel prize (*Science*, 6 October 2006, p. 34). Now, a California research team has made the controversial claim that such RNAs can have the opposite effect: They can turn genes on.

This surprising skill—dubbed RNAa, because the RNAs activate genes—is described this week in the online edition of the *Proceedings of the National Academy of Sciences*. If the claim is sustained, RNAa would be a powerful biological tool and could lead to new therapies for diseases such as cancer. But some scientists say the results may reflect an indirect outcome of RNAi, rather than a new way to activate genes. "It's going to be a question of whether this holds up," says Erik Sontheimer, an RNA researcher at Northwestern University in Evanston, Illinois.



New phenomenon? Compared to typical prostate cancer cells (*bottom*), ones administered a short double-stranded RNA (*top*) boost production of a protein (green) encoded by a tumor suppressor gene.

CREDIT: LONG-CHENG LI

RNAi is generally thought to thwart gene translation—the double-strand RNAs cut up a gene's mRNA or block its ability to make protein. But in lower organisms, it can also work at the level of transcription, preventing a gene from even making its mRNA. Long-Cheng Li, a postdoc in the lab of cancer researcher Rajvir Dahiya at the University of California, San Francisco (UCSF), tried to use RNAi to block transcription of the human E-cadherin tumor suppressor gene. When Li added synthetic RNAs that specifically targeted the gene's DNA sequence to human prostate cancer cells, E-cadherin levels unexpectedly went up, not down. "It was immediately quite obvious," Li recalls.

Li then used synthetic RNAs to boost expression of two other genes in cultured cells and now says he can activate numerous tumor suppressor genes with RNAa. If the effect turns out to be predictable, RNAa "could be very powerful, in terms of potential [anticancer] therapeutic application," says John Rossi, an RNA expert at the City of Hope National Medical Center in Duarte, California. Although not every gene is susceptible to RNAa, Li says he's mostly worked out rules for activating those genes that are. He plans to make these rules "readily available to the public" after ironing them out and activating more genes. UCSF has filed for a patent on RNAa.

One key question is whether Li's RNAs are activating genes by silencing others, which would just be RNAi by another name. For example, proteins called negative transcription factors can prevent genes from being transcribed; silencing the genes for these proteins could activate genes they control. Although the UCSF group has not found evidence that this is happening, "formally, that's still a possibility," says Rossi.

No one yet knows how small RNAs could turn genes on, especially for so long. RNAi typically silences genes for 5 to 7 days, but RNAa boosted gene activity for up to 13 days. The molecular machinery underlying RNAi appears to be involved in RNAa, raising the question of how the same enzymes can sometimes turn genes off, and sometimes on. "What makes one siRNA [small interfering RNA] a silencer, and what makes the other one an activator?" asks Sontheimer. "No clue."

Sontheimer also wonders why other groups haven't seen similar gene activation, especially in microarray studies of RNAi that examine thousands of genes. At least four groups have now reported that siRNAs are gene silencers at the level of transcription in mammals, but none have seen gene activation. One of the groups even silenced the gene for E-cadherin, the same one that ▶

Rice Krispies

NEW DELHI—Indian activists have torched the first field trial of a genetically modified food crop. Genetically modified cotton is widely grown in India, but last week, a small field trial of hybrid Bt Rice genetically modified for insect resistance was burnt to ashes at Rampur village in Haryana. It was one of 12 field locations belonging to Maharashtra Hybrid Seeds Company Limited (MAHYCO), Mumbai.

Officials with MAHYCO, owned in part by global seed giant Monsanto, say about 200 activists belonging to the farmers' Bhartiya Kisan Union forced their way into the controlled plot and shouted anti-GM slogans before torching the plot, which was ready to be harvested. Rakesh Tikait, a leader of the group, which is one of several of Indian farmers' groups, told *The Indian Express* that such trials would contaminate the soil and affect yield from existing varieties. "The crop was being grown in isolation as per the [rules], following all safety measures," responds MAHYCO general manager Mahendra Kumar Sharma, who called the attack "deplorable."

Late last month, the nation's Supreme Court put a moratorium on new approvals of genetically modified field releases, and officials must now respond to complaints by activists that permissions had been granted "recklessly." A hearing on the matter is expected next month. **–PALLAVA BAGLA**

Jockeying Planetary Missions

NASA's science budget is tight, but the agency nevertheless approved work on three planetary science proposals—to examine Venus's atmosphere, probe the moon's interior, and return an asteroid sample. Each team gets $1.2 million to provide a more detailed plan for a mission which must cost less than $425 million; the winner will be chosen next year once the studies are complete.

The agency also plans to continue at least one of two missions now in flight. One option would be to redirect the Deep Impact spacecraft that visited Comet Tempel 1 in 2005 to Comet Boethin, to compare the two objects. The other choices would be to focus a camera from the same spacecraft on possible Earth-sized planets around stars, or to send the Stardust spacecraft, to check on changes to Tempel 1 since its encounter with Deep Impact.

"One of the great surprises of comet explorations has been the wide diversity among the different cometary surfaces imaged to date," says Michael A'Hearn, the University of Maryland astronomer who would lead the Boethin mission. **–ANDREW LAWLER**

UCSF turned on. "There's really no indication yet as to why they [at UCSF] see the exact opposite thing," says Sontheimer.

But Rossi—who co-authored one of the silencing papers—says it's possible that he and others missed RNAa because they didn't expect it. "We never did look for upregulation," he admits. And Steve Baylin and Angela Ling, the Johns Hopkins University researchers who silenced the E-cadherin gene with siRNA, find the UCSF report credible. "I'm not sure there's any conflict in the data," says Baylin, who points out that the RNA used by the UCSF group targeted a different part of the gene's sequence from the ones his group employed. "[Gene] region may be the real key."

Fred Gage, a neuroscientist at the Salk Institute for Biological Studies in San Diego, California, calls the UCSF results "intriguing." Two years ago, Gage found a short double-stranded RNA in adult neural stem cells that can activate genes important for neuron function. Gage's activating RNA was naturally made by the cells, while Li used synthetic RNAs. If the UCSF group found similar RNAs in natural systems, that "would take this to another level," Gage said. Li says he now has some evidence for that.

If RNAa is indeed a new phenomenon, researchers trying to exploit RNAi will need to avoid activating other genes beyond the one they're trying to silence, an "off-target" effect that could hamper research applications and new therapies (*Science*, 12 November 2004, p. 1124). But if it does occur naturally, RNAa could provide new insights into gene regulation, adding yet another surprising role to RNA, the molecule of the moment. "If this holds up," says Sontheimer, "it seems there's no end to the number of regulatory mechanisms that small RNAs can access."

–KEN GARBER

Ken Garber is a freelance writer in Ann Arbor, Michigan.

AVIAN INFLUENZA

New H5N1 Strain Emerges in Southern China

A troubling new strain of H5N1 avian influenza has emerged in China over the past year. The group that identified the virus warns that it may be resistant to current poultry vaccines and is possibly now spreading a third wave of bird flu infection across Asia.

International animal health authorities are taking notice but not panicking yet. The emergence of a new, genetically distinct strain "is cause for concern," says Peter Roeder, a virologist with the United Nations' Food and Agriculture Organization (FAO) in Rome. But he adds that claims about its resistance to vaccines "need clarification to justify the conclusions."

Yi Guan, director of the State Key Laboratory of Emerging Infectious Diseases at University of Hong Kong, along with colleagues there and at St. Jude Children's Research Hospital in Memphis, Tennessee, report their findings online this week in the *Proceedings of the National Academy of Sciences*; the paper will appear in the 7 November print edition.

Surveillance. By sampling poultry in markets in southern China, Yi Guan (*center*) and colleagues spotted a new strain of the H5N1 virus.

CREDIT: STATE KEY LABORATORY OF EMERGING INFECTIOUS DISEASES AT UNIVERSITY OF HONG KONG

Guan and his colleagues identified the new strain and a general upswing in overall H5N1 infections through their ongoing surveillance of poultry markets in six provinces of southern China. The team found that from July 2005 through June 2006, the percentage of ducks, geese, and chickens infected with H5N1 climbed to 2.4% of those sampled, up from 0.9% the previous year. The findings suggest the virus remains firmly entrenched in the region, particularly among domestic ducks and geese.

They also found that a new dominant strain had emerged. This H5N1 sublineage, which they call the Fujian strain, was first detected in March 2005 but turned up in only one sample from July to September that year. However, the Fujian strain accounted for 95% of all samples collected from April to June 2006. Several other strains previously circulating in the region dropped off the radar. "It appears that [previous] sublineages have been replaced by this new variant," Guan says.

The researchers found that the hemagglutinin gene from recent human cases reported in China also belonged to the Fujian strain, confirming that it does infect humans. Fujian-like strains were also isolated by other surveillance efforts in Hong Kong, Laos, and Malaysia, indicating it is already spreading beyond southern China.

To check the effectiveness of current vaccines, the group screened blood sera collected from chickens to identify samples from vaccinated animals. They then tested how well 76 of those samples selected at random neutralized three viruses, including the new Fujian strain. Most samples neutralized the older virus strains but had minimal effect on the Fujian strain.

Guan and his colleagues speculate that the new virus may be resistant to current vaccines and that it may have emerged in response to the widespread poultry vaccination in southern China. "Our data show a need to change [currently used] vaccines," Guan says.

Other researchers praise the surveillance effort for spotting the new H5N1 strain. But they are more cautious about the implications for vaccines. Les Sims, a veterinarian based in Manunda, Australia, who advises the FAO on poultry vaccination programs, says, "We recognize that the use of vaccination has the potential for driving antigenic change in these viruses." But he notes that different strains of H5N1 emerged and became dominant even before there was widespread use of vaccines. To demonstrate conclusively that current vaccines aren't working, researchers would need to vaccinate live chickens, infect them with the new strain, and observe the results, Sims adds. Guan agrees and says they are now planning just such an experiment.

Another point on which the two agree is the need to continue postvaccination surveillance efforts—such as Guan's in southern China—to spot and deal with any vaccine-resistant strains that do emerge. **–DENNIS NORMILE**

BIOTERRORISM AGENTS

U.S. Panel Wants Security Rules Applied to Genomes, Not Pathogens

A U.S. government panel says that advances in synthesizing genomes are outpacing the country's attempt to prevent bioterrorism. The solution, says the National Science Advisory Board for Biosecurity (NSABB) in recommendations adopted last week, is to regulate potentially dangerous gene sequences instead of a list of known pathogens.

The board was set up 2 years ago by the Department of Health and Human Services (HHS) to help develop safeguards against the accidental and deliberate misapplication of life sciences research. But researchers can now engineer biological agents that are functionally similar to pathogenic microbes and yet fall outside the scope of rules governing their handling, scientists explained at a board meeting last week. The rules, established after the anthrax attacks in the fall of 2001, mandate special security procedures for the handling and shipping of approximately 100 so-called select agents—microbes, viruses, and toxins that the government views as potential bioterrorism threats.

"The current rules apply only to biological entities whose nucleic acid sequences are identical to those of agents listed by the government," says Stanford microbiologist David Relman, chair of NSABB's working group on synthetic genomics. "But what about a genetic variant of a select agent that still exhibits the same properties as the agent? And what about novel pathogens that can be engineered using combined genetic material of multiple select agents?"

To deal with those scenarios, says the board, the government needs a "framework based on predicted features and properties encoded by nucleic acids" instead of "the current finite list of specific agents and taxonomic definitions." Says Michael Stebbins of the Federation of American Scientists in Washington, D.C.: "What they're saying is that the government needs to stop thinking about genomics in terms of organisms and start thinking about it in terms of DNA content."

That approach may assist government regulators, agrees Gigi Kwik Grönvall, a biosecurity expert with the University of Pittsburgh, Pennsylvania. But a regulatory framework based on properties of gene sequences "may not provide the clarity needed for the person at the lab bench trying to make sure he or she does not run afoul of the law," she says. "What that person needs are clearly defined dos and don'ts, not complicated algorithms."

In another recommendation, the board calls for repealing a 2-year-old law that bans the synthesis of the smallpox virus (*Science*, 11 March 2005, p. 1540). The prohibition applies to "any derivative of the variola major virus that contains more than 85% of the gene sequence," a definition that covers several pox viruses commonly used by researchers, including a strain used for making vaccines. But Ed Hammond of the Sunshine Project, a bioweapons watchdog group in Austin, Texas, warns that repeal "could result in a proliferation of the virus and its parts."

The board also wants the government to require companies to screen orders for synthetic DNA against the genomes of select agents and to maintain a record of purchase orders. Neither procedure is currently mandated by law. **–YUDHIJIT BHATTACHARJEE**

Safe handling. Special security procedures exist for working with certain pathogens.

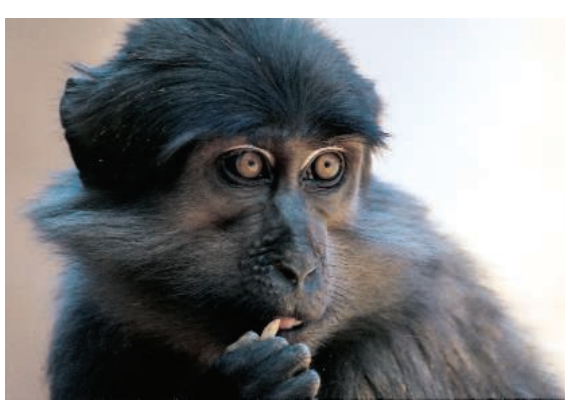
AIDS RESEARCH

Center Puts Hold on Mangabey Experiments

In a letter made public last week, Yerkes National Primate Research Center in Atlanta, Georgia, withdrew a request to conduct experiments with sooty mangabey monkeys that could unravel fundamental riddles about how HIV causes AIDS. The U.S. Fish and Wildlife Service (FWS) considers the primates an endangered species, so Yerkes's proposal had attracted intense criticism. Yerkes says it hasn't abandoned plans for such research, however. It's waiting for FWS to reassess whether the sooty mangabey is truly endangered.

AIDS researchers study sooty mangabeys because SIV, HIV's cousin, naturally infects these African monkeys but rarely causes harm. Yerkes has more than 200 sooty mangabeys, the largest captive colony in the world. FWS has long granted Yerkes a permit to collect blood from the animals and perform limited biopsies for research. In January, as part of the permit renewal process, Yerkes requested a "variance" that would allow the institution for the next 5 years to cause disease or euthanize up to 20 animals annually to further AIDS research.

Yerkes asked for the variance in part because of increasing interest in why sooty mangabeys have high levels of SIV in their blood but show no immune damage, unlike the rhesus macaques that AIDS researchers more commonly study. "It's such an important question," says Guido Silvestri, a pathologist who in February moved from the primate center to the University of Pennsylvania in Philadelphia. Silvestri's work suggests that mangabeys remain unharmed by SIV because, unlike humans, they do not "overactivate" their immune systems when confronted with the virus. "There are a lot of studies we could do" if ▶



Truly endangered? A Georgia research center says no, and wants more freedom to do invasive tests with its sooty mangabey colony.

CREDITS (TOP TO BOTTOM): NIAID/NIH; YERKES NATIONAL PRIMATE RESEARCH CENTER

the variance were granted, he says, including overactivating mangabeys' immune systems to see whether that causes AIDS.

When FWS invited public comment on Yerkes's request in May, opposition surfaced, including a letter from primatologist Jane Goodall and 18 other scientists. A key point of contention: The proposal noted that Yerkes funds a conservation effort for sooty mangabeys in Côte d'Ivoire. Goodall and her co-authors warned that "Approving Yerkes's application could open the floodgates to future permit applications premised on allowing entities to kill or otherwise harm endangered species in exchange for making contributions to conservation programs."

Jim Else, Yerkes associate director for research resources, challenges this idea of a quid pro quo. "It wasn't 'Give us this, and then we'll do that,' " says Else, noting that FWS encourages permit applicants to explain how they are helping species in the wild. "We were already providing the support to conservation."

More important, Else says, FWS wrongly classifies sooty mangabeys as endangered because it relies on an old taxonomy that lumped species and subspecies together. "The taxonomy has changed beyond all recognition," says Else, a veterinarian.

Even some leading conservationists support this contention. The World Conservation Union, which publishes a "red list of threatened species," considers sooty mangabeys—*Cercocebus atys atys*—as "near threatened," two notches down from endangered. However, *Cercocebus atys lunulatus*, or white-naped mangabeys, are at the top of the endangered list. FWS makes no such distinction, listing all mangabeys as yet another species, the red-capped *Cercocebus torquatus*. The sooty mangabey "is not as threatened as people think it is," concludes Anthony Rylands, deputy chair of the primate specialist group for the red list.

Michael Kreger, who works in the FWS branch that oversees foreign species on the endangered list, says the agency currently is reviewing the status of the sooty mangabey. In its 18 September letter, first reported by the Associated Press, Yerkes wrote FWS that it wanted to withdraw its variance request "in light of the possible reconsideration of the sooty mangabey classification status."

–JON COHEN

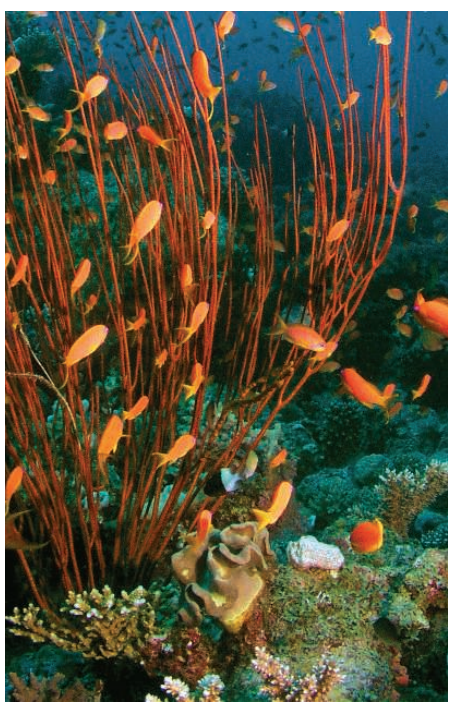
ECOLOGY

Global Loss of Biodiversity Harming Ocean Bounty

Environmental groups often argue that biodiversity offers tangible benefits to people. Now, a group of ecologists has put that argument to the test with the most comprehensive look yet at the human impact of declining marine biodiversity. On page 787, they report that the loss of ocean populations and species has been accompanied by plummeting catches of wild fish, declines in water quality, and other costly losses. They even project that all commercial fish and seafood species will collapse by 2048. "It's a gloomy picture," says lead author Boris Worm of Dalhousie University in Halifax, Canada. Yet the team provides a glimmer of hope, concluding that people still have time to recoup these ecosystem benefits if they restore biodiversity.

Although none of these points is new, some experts say the study strengthens the case for the practical value of biodiversity by marshaling multiple lines of evidence and taking a global look. "This is a landmark paper," says Jane Lubchenco of Oregon State University in Corvallis. Others aren't convinced yet. "It falls short of demonstrating that biodiversity losses are the primary drivers of why the services have declined," says Donald Boesch of the University of Maryland Center for Environmental Science in Cambridge.

Past studies of so-called ecosystem services have demonstrated, for example, that a rich array of pollinators creates greater yields for coffee farmers (*Science*, 20 August 2004, p. 1100). But proving that such benefits exist on a global scale has been difficult, particularly for the oceans, which remain poorly studied.



CREDIT: AMIR GUR

At your service. Highly diverse ecosystems, such as the Red Sea, provide many more ecological services than species-poor ecosystems.

To gauge whether the loss of marine biodiversity matters, Worm and his co-authors reviewed all the data they could find on the issue. They discovered a consistent pattern. In 32 small-scale experiments, higher diversity of either marine plants or herbivores led to benefits such as greater ecosystem stability and 80% more biomass. A review of 12 estuaries and other coastal ecosystems found the same trend. Those with more species had lower rates of collapse of valuable fisheries than systems that were relatively species-poor to begin with. The team also argues that loss of filter feeders led to a decline in water quality, including depletion of oxygen, in regions such as the Chesapeake Bay.

Data for 64 large marine ecosystems showed that fisheries are collapsing at a higher rate in species-poor ecosystems than in species-rich ecosystems. "Within my lifetime, I might see global cessation of wild fisheries," Worm says. The good news is that closing fisheries and establishing protected areas boosted the number of species in these regions by 23% on average and increased catch-per-unit effort four-fold in nearby waters, although overall yield didn't increase much.

Still, Boesch and others note that it's difficult to prove that loss of diversity causes the decline in services. Boesch says that in the Chesapeake Bay, factors such as excessive fertilizer runoff probably are the real cause of the decline in water quality. Ray Hilborn, who studies fisheries at the University of Washington, Seattle, adds that fishing doesn't necessarily causes ecosystems to be less productive; the long-exploited Mediterranean, he points out, continues to be productive.

Worm and his colleagues call for the creation of new marine reserves, sustainable management of fishing, and tighter control of pollution. Those are well-worn recommendations, but Worm says the team's analysis of the consequences of not taking action, especially the loss of wild fisheries, gives them greater weight. "If you can see the bottom of the barrel, that changes things."

–ERIK STOKSTAD

Wolves, bears, and other large carnivores are returning to western Europe. But is there still room for them?

The Carnivore Comeback

ARBAS, FRANCE, AND MARIAZELL, AUSTRIA— This used to be just another sleepy village in the Pyrenees. But lately, the mayor of Arbas, population 250, has received death threats, the quiet central square has been turned into a battlefield between protestors and police, and bottles of sheep blood have been smashed against the sandstone facade of the town hall.

Arbas has become the epicenter of one of France's most hotly debated ecological issues: the government's plan to save the remaining brown bear population in the Pyrenees by reintroducing animals captured in Slovenia, where they are still abundant. Arbas's mayor, François Arcangeli, enthusiastically endorses the plan, and he chairs Pays de l'Ours-ADET, a nonprofit organization promoting peaceful coexistence between bears and humans. So when the government picked sites near Arbas to release three Slovenian bears earlier this year, it was hoping for little resistance; instead, Arbas has become a magnet for frustrated opponents, primarily sheep farmers who say their livelihoods are threatened.

France's battle of the bears is one of the most vicious examples of a struggle taking place in several European countries. The original populations of bears, wolves, lynx, and wolverines—the four main large predators native to Europe—were exterminated from many of the western countries in the 18th and 19th centuries as habitat disappeared and hunters sought out the last of the hated predators. But in recent decades, carnivores have been making a comeback, increasing in numbers and expanding their territory.

They have often done so with little or no human help. Bears, wolves, and lynx naturally travel hundreds of kilometers in search of food and mates, and the dismantling of border fences between western and eastern Europe has allowed new immigration from the often-robust populations in former communist countries. In some cases, governments have urged the process along by transplanting animals from eastern Europe.

The comeback has triggered a wave of new research into the behavior and population dynamics of large carnivores. Scientists are studying how many individuals are needed to sustain a viable population, for instance, and what the most effective management strategies are. They are tracking how far the animals wander, who mates with whom, and how barriers such as highways affect both migrations and genetic diversity.

But although a science-based management plan is essential if the animals are going to thrive, that alone is not sufficient, experts agree. The overriding question, they say, is whether citizens of these densely populated and highly developed countries will be willing to coexist with the animals—even if they occasionally devour livestock and scare unsuspecting humans. The key to success, says John Linnell

Wild things. Lynx, wolverines, and wolves are increasing in numbers and in territory across western Europe.

CREDITS (TOP): O. ALAMANY AND E. VICENS/CORBIS; (BOTTOM, LEFT TO RIGHT) GUY EDWARDES/GETTY; GETTY IMAGES; JOHN LINNELL

Wary welcome. Brown bears like this one in the Pyranees in France have sparked vigorous debate.

of the Norwegian Institute for Nature Research in Trondheim, Norway, "is to get people used to the idea of having something in their backyards that is wild and a little out of control."

A wild hope

For centuries in Europe, big carnivores were seen as dangerous and shrewd enemies, and killing them was considered a virtue. But in the 1960s and 1970s, as biodiversity rose on the political agenda, conservationists and governments across western Europe began rallying support for new policies to protect the dwindling populations. Supporters concede that in western Europe, big carnivores aren't needed to sustain a healthy ecosystem; hunters are usually happy to keep populations of prey animals such as deer and wild boar in check. But the photogenic animals can act as "umbrella species": The decision to protect their large habitats often results in a whole series of measures—such as restricting development and building migration corridors over highways—that will help protect many other, less charismatic species.

Carnivore supporters offer a moral argument as well. True, with the exception of the Iberian lynx (see sidebar, p. 749), none of Europe's big carnivores is endangered—in fact, they are thriving in large parts of eastern Europe. But eastern countries shouldn't bear the burden of conservation alone, argues Olivier Hernandez of the French WWF, formerly the World Wildlife Fund. "We also maintain the Louvre, even though there are great museums in eastern Europe," he says. Nor should rich countries such as France and Austria preach about conservation in the developing world if they can't sustain their own carnivore populations, says bear expert Beate Striebel of WWF in Austria. "Elephants cause much more damage and are more dangerous than bears," she says.

Although it is still early days, conservationists say, there is reason for optimism. Wolves have returned to Sweden, where they now number about 100, and to Germany, where more than a dozen have taken up residence in a military training ground on the Polish border. Small populations of reintroduced lynx have gained footholds in Switzerland, eastern France, and southwestern Germany, and natural immigrants are thriving in southern Sweden. In northern Scandinavia, populations of wolverines are small but stable or even increasing. Bear populations are also small but stable in Austria and Italy, and the one in the Pyrenees, although still hanging in the balance, may just make it. "If you look at Europe as a continent, we shouldn't complain," says ecologist Luigi Boitani of the University of Rome "La Sapienza."

Save our sheep. Shepherds in southern France protest the release of Slovenian bears to boost the dwindling local population. The banner reads, "Freedom for bears, danger for people."

Room to roam

As they search for the best ways to support these often-fragmented populations, scientists are gathering more precise data on them. So far, even basic population estimates have largely been based on extrapolations and guesswork. Now, genetic tools are providing a far more accurate tally and also providing new insights into how the animals use their space.

Too close for comfort. Bruno, a brown bear that found its way to Germany in May, had developed a troubling taste for lambs and other livestock.

In Austria and France, genotyping of hair and scat has enabled officials to trace damage reports to specific animals so they can better determine whether a single "problem bear" needs to be targeted for tracking or possible interventions. In Austria, DNA evidence suggests that the bear population numbers just 20—and not the 25 to 30 previously estimated—despite the births of 27 cubs between 1991 and 2005. Such studies have also yielded worrying signs of inbreeding. In one region, a single male fathered all 12 cubs born between 1994 and 2003, including litters with two of his daughters.

Using Global Positioning System–enabled radio collars, scientists are learning more about migration patterns. Radio collars can also help scientists determine where to put "green bridges" to allow animals to cross large highways safely. One radio-tagged wolf migrated more than 300 kilometers from Parma, Italy, to Nice, France, for instance, whereas a bear was spotted leaving the Pyrenees and approaching the Toulouse suburbs, 50 kilometers to the northeast. (It was eventually captured and returned to the mountains.)

The animals' surprising mobility highlights one acute problem in protecting them. In most of Europe, wildlife management is the responsibility of a patchwork of organizations: In different areas, the agriculture ministry, the environment ministry, or even hunting organizations have formal responsibility for local management of large carnivores. Now, several ecologists are working

CREDITS (TOP TO BOTTOM): ASSOCIATED PRESS; JOHANNES SIMON/AFP/GETTY IMAGES

with the European Union to develop a population-based plan that recognizes that borders mean little to such animals. The new plan would take into account the genetic diversity of the populations and possible corridors among them and will attempt to draw up rules that, if not the same from region to region, at least don't actively conflict with one another.

Good neighbors?

That still leaves one major obstacle, however: overcoming public opposition. "I hate to admit it as an ecologist, but the most pressing issues are related to social science," Linnell says. "Understanding the sociology of coexistence is really the key."

The problem was painfully illustrated by the fate of Bruno, as the media called him—a bear born in Italy that crossed Austria and finally ended up in southern Germany last summer. The first wild bear to set foot in the country in nearly 100 years, Bruno was warmly welcomed; Bavarian state environment minister Werner Schnappauf even held a press conference to celebrate his arrival. But those feelings cooled when Bruno's taste for sheep, chickens, and caged rabbits—and his apparent fearlessness of humans—became evident. After weeks of fruitless attempts to capture him, he was summarily shot by hunters commissioned by the Bavarian government.

Worries about carnivores ravaging livestock and putting humans in danger have triggered opposition to their recent expansions throughout Europe, and especially where they have been reintroduced. Sheep farmers in the Pyrenees say that the five bears released so far this year threaten their livelihoods and create a mortal danger for shepherds, hikers, and hunters. Mountain guide and former shepherd Louis Dollo, a vocal spokesperson for the antibear movement, says the program was forced on the

DATA SOURCE: LCIE REPORT FOR THE UNEP-WCMC PROJECT, BY THE LARGE CARNIVORE INITIATIVE FOR EUROPE

fiercely independent region by conservationists and bureaucrats in Paris. "These people don't have a clue about life in the mountains," he says.

Tensions in the region have escalated so badly recently that when Palouma, a female brown bear released in April, plunged from a cliff late August and died, some suspected foul play. (An official investigation into her death is ongoing.) Ecologist Pierre-Yves Quenette, head of the government team that releases bears and studies them afterward, says the recriminations and threats have become so intense that he had to take time off earlier this year to preserve his sanity.

Geographer Farid Benhammou, a reintroduction supporter who is working on a Ph.D. thesis about the battle, says the fierce resistance stems in part from broader discontent among farmers about the troubled economy and the influx of urban people into rural areas. "The bears have become a scapegoat for everything that's wrong," he says.

That doesn't mean that bears are problem-free, however. Although bear supporters maintain that the risk to humans is greatly exaggerated—no human being is known to have been killed by bears in the Pyrenees for at least 150 years—they concede that the damage to livestock is real. Bears kill some 200 sheep annually in the Pyrenees alone. Wolves and lynx cause damage throughout Europe, especially in areas where they are newcomers and farmers haven't adapted to their presence.

The French government is trying to find a solution by compensating farmers for lost sheep, giving them the benefit of the doubt when a bear attack is suspected but not proven. It also sponsors the construction of mountain huts for shepherds (until recently, most sheep wandered around unguarded) and offers farmers subsidies to get a trained dog to help ward off attacks. But farmers say the compensation isn't enough, and most wouldn't shed a tear if the entire bear population dwindled to zero, Dollo concedes.

Proponents of the reintroductions, meanwhile, are trying to play into the popularity of bears in the general population. They launched a special cheese, for instance, imprinted with a bear paw, that only farmers committed to protecting bears can produce. And Alain Reynes, director of Pays de l'Ours-ADET, argues that bears will lure, not deter, tourists, noting that the Italian region of Abruzzo has seen tourism increase after it started billing itself as bear and wolf country. (That the average hiker or mountain biker is extremely unlikely to see a bear appears to be irrelevant.) Ecologists and advocates across Europe are also working to woo the support of hunting groups, which wield significant power.

Carnivore advocates say that western Europe as a whole could take some lessons from Austria and Italy. After considerable ups and downs, both countries have learned anew to live with bears. After a particularly bad run of bear damage in 1994, Austria hired four "bear advocates," biologists who are responsible for assessing damage and working with local residents, helping them to bear-proof farms and hunting stations, and explaining how to handle encounters with bears.

That experience will need to be replicated if the species are to remain in their reconquered territory, says Linnell. "It's not about having these animals in a national park," he says. No park in Europe can sustain even a remnant population. "We want to get people to accept that wolves and bears are part of the modern 21st century landscape."

–MARTIN ENSERINK AND GRETCHEN VOGEL

CREDIT: REUTERS

Precious few. If the Iberian Lynx doesn't survive, it would be the first documented feline extinction since the saber-toothed tiger.

On the Brink

Three of the most genetically valuable kittens on earth were born in southern Spain last year. The first captive-born members of the world's most endangered feline species, the Iberian lynx, the cubs represent a ray of hope in an otherwise grim story. Although the lynx was once prevalent from the Pyrenees south to the Mediterranean, now only 200 individuals are left, surviving in two fragile pockets in Andalucia.

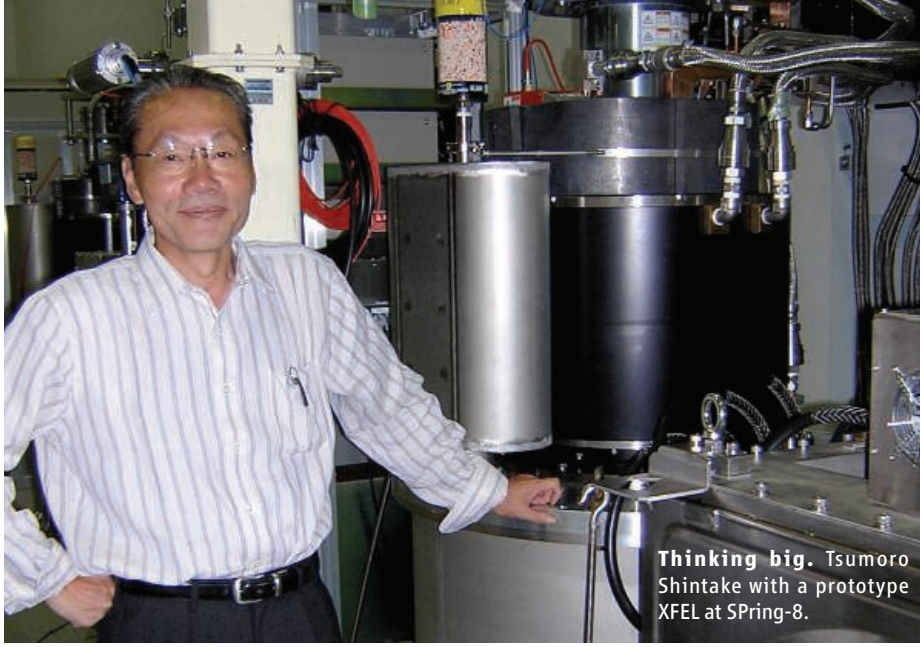
The Iberian lynx is about twice the size of a housecat and half the size of the more common Eurasian lynx, which is making a comeback elsewhere in western Europe (see main story). The Iberian population was small but sustainable in the early 1980s with about 1100 animals. But it was devastated by an outbreak of two exotic diseases that killed up to 90% of the region's wild rabbits—the lynx's primary prey. At the same time, Spain and Portugal, as new members of the European Union, received an influux of funding for new roads, high-speed trains, and tourism infrastructure, squeezing the lynx's habitat.

"It was a huge emergency situation," says Astrid Vargas, who now heads the Program for Ex-situ Conservation of the Iberian Lynx, based in the Doñana National Park. Last-ditch efforts to protect habitat and rebuild the rabbit population seem to have helped: One of the populations is stable, and the other has grown slightly since 2002. But the animals are still on the brink, and a fire or epidemic could quickly wipe out the remaining survivors, Vargas says.

The captive breeding program Vargas heads is designed to release animals into currently lynx-free areas by 2010. Now in its second year, the program has produced nine cubs, five of which have survived. Along the way, Vargas and her colleagues are collecting a wealth of data about the animals' behavior and reproduction. One of the most important lessons was that young cubs go through an extremely aggressive phase a few months after birth, fighting so brutally with their littermates that they often kill each other. After losing one of the first three cubs in such a fight, the scientists now separate the young animals for a few critical weeks.

But most crucial, say Vargas and others, is the search for an appropriate spot to release the animals. Scientists are seeking 10,000 hectares of habitat with healthy rabbit populations and minimal roads—seven lynx have been killed in road accidents in the last 18 months. That's not easy to find, Vargas says, but is the only way the animal will survive. "Captive breeding … is not a salvation for the lynx. If we're breeding but there is no habitat, we're not saving the species."

–G.V.



Thinking big. Tsumoro Shintake with a prototype XFEL at SPring-8.

CREDIT: D. NORMILE/*SCIENCE*

MATERIALS SCIENCE

Japanese Latecomer Joins Race To Build a Hard X-ray Laser

X-ray free-electron lasers are the next big thing in high-energy probes of matter. With U.S. and European machines in the works, Japan wants into the club

SAYO, HYOGO PREFECTURE, JAPAN—It's the scientific version of keeping up with the Joneses. Once researchers in one region plan a big, new experimental device, researchers everywhere want their own. The latest example: x-ray free-electron lasers (XFELs), which promise beams that are vastly brighter and with higher energy and shorter pulses than today's workhorse synchrotron x-rays.

These "hard" x-ray wavelengths—down to 0.1 nanometer—promise to reveal the structure of proteins that have eluded other techniques and nanometer-scale features in materials. Pulses as short as 100 femtoseconds or less will act as strobes to produce movies of molecular bonds breaking and forming in chemical reactions. And astrophysicists will become experimentalists, using beams 10 billion times brighter than synchrotron radiation to create the extreme state of matter believed to exist within forming stars.

And that could be just the beginning. "I expect to be surprised by scientific opportunities we are not even talking about now," says John Galayda, head of XFEL development at Stanford Linear Accelerator Center (SLAC) in Menlo Park, California, which last month broke ground on its Linac Coherent Light Source (LCLS).

With breathtaking science at stake, groups in Japan, here at the RIKEN Harima Institute, and in Europe are also rushing to bring XFELs on line. "I wouldn't call it a race, but with such broad interest for science, it is no surprise that [researchers] in three regions of the world want to have a facility of their own," says Reinhard Brinkmann, who leads the European effort based at the German Electron Synchrotron (DESY) research center in Hamburg. "Free-electron lasers are amazing things which herald a new era in photon science," says Janos Hajdu, a synchrotron radiation specialist at Uppsala University in Sweden.

XFELs rely on new approaches to generating both x-rays and laser light. Current synchrotrons send electrons whizzing around a storage ring a kilometer or more in circumference. As the electrons pass through bending magnets or "wigglers" that curve their path, they throw off photons at soft x-ray wavelengths.

Instead, XFELs have everything in a line: an electron source, a linear accelerator that propels the electrons, and an undulator, which has two rows of magnets of alternating polarity that make the electrons zigzag up and down as if on some magnetic slalom course. Just as a skier rounding a gate throws a spray of snow down the slope, the electrons throw forward a clutch of photons with each zig and zag. The interplay between the electrons and the photons produces an x-ray laser thanks to a phenomenon called self-amplification of spontaneous emission, or SASE (*Science*, 10 May 2002, p. 1008). Laser light is coherent, meaning that all the photons are in phase, or oscillating in lockstep—a quality missing from synchrotron light.

Although all three planned systems share the same basic setup, subtle differences give each of them strengths and weaknesses. "The final targets of the XFEL projects are the same, but the means are different," says Tsumoru Shintake, who heads accelerator development for Japan's XFEL.

The first project to come online will be Stanford's LCLS. Much of the key research underpinning XFELs was done at SLAC beginning in the early 1990s. And SLAC got a head start by using a 1-kilometer stretch of its now-idled linear accelerator, or linac. The SLAC group estimates that reusing its linac has saved more than $300 million, giving a total construction cost of $379 million. LCLS will have one undulator providing hard and soft x-rays to up to six experimental stations. Galayda says the group expects to generate its first x-rays by July 2008 and to start experiments by March 2009.

Japan's entry is the SPring-8 Compact SASE Source (SCSS), just now getting under construction here. Latecomers to the field, the team is using some homegrown technology to cut cost and size. "We're taking the first step toward making XFELs smaller and cheaper so more [institutions] can consider developing their own," boasts SCSS project leader Tetsuya Ishikawa. Whereas the other two machines will generate electrons by firing a laser at a metal target, the SCSS heats a cathode to produce electrons. Eliminating the laser simplifies the system but requires careful compression of the cloud of electrons before they go into the linac.

The wavelength of the output x-rays is a tradeoff between the energy of the electrons and the undulator period. The Americans and Europeans have opted for higher electron energies and longer periods. The Japanese placed their bets on the opposite approach. The SCSS's linac produces lower energy electrons, but then its undulator magnets are placed inside the vacuum tube housing the electron beam, allowing the gap between the magnets to be a slim 3 to 4 millimeters. In the U.S. and European machines, the undulator magnets are outside the vacuum tube, so they must be farther apart.

Although this arrangement sounds simple, Hajdu says it required technological advances in controlling the accelerating electrons and in the precision of the undulator. "There was huge skepticism in the community about [the Japanese approach] early on," Hajdu says. But in August, at an XFEL conference in Berlin, Shintake reported that a prototype machine incorporating SCSS's new technologies had successfully produced a beam.

The simplicity and compact size of the SCSS result in a construction cost of $315 million, although it is not directly comparable to the costs of the other projects as it excludes personnel, administrative, and instrument costs. The Japanese team also came a long way in a short time, starting on development just 2 years ago. "We are beginners in this community," Shintake says. He says his group will complete the project by March 2010 and will start experiments at two stations the following year.

The European XFEL is "a much more grandiose system," Hajdu says. It got that way partly by accident. The European XFEL was originally packaged with the proposed TESLA particle physics project, which called for a superconducting linac. TESLA was abandoned, but XFEL development continued. "It would have been stupid not to use [the superconducting technology]," says Brinkmann.

So whereas the other systems use conventional linacs, the European XFEL will have a 1.6-kilometer-long superconducting accelerator capable of supplying electrons to three hard and two soft x-ray beamlines supporting 10 experimental stations. This power and flexibility makes the European XFEL the priciest system, at $1.1 billion. Brinkmann says the Europeans are now engaged in the "nontrivial" exercise of finalizing funding among contributing countries. Germany will pay about 60% of the bill, with other European countries, Russia, and China contributing the rest. He expects a final go-ahead in early 2007, with experiments starting in 2013.

All three machines are aiming for "hard" x-rays down to about 0.1-nanometer wavelength with pulse durations of 100 femtoseconds and a trillion photons in each pulse. But the number and pattern of the pulses differs significantly. SCSS and LCLS will typically put out single pulses of light at a rate of 60 and 120 pulses per second, respectively; the European XFEL will put out bundles of up to 3000 pulses 10 times per second—a machine gun to the other two pump-action rifles.

Massimo Altarelli, a theoretical physicist who is the European Union team leader for the project, says the 30,000 pulses per second will allow a much more rapid accumulation of data, a particular advantage for "pump and probe" experiments in which an initial pulse induces a photochemical reaction or creates warm, dense matter from a solid target and a second pulse examines the changes a few hundred femtoseconds later. Such experiments must be repeated thousands of times to accumulate statistically significant amounts of data. "If you have thousands of pulses per second, there is a substantial advantage," Altarelli says.

But there is a catch: Observing all the signals produced requires dramatic advances in detectors. "With today's instrumentation, you're not going to be able to really take advantage of these features," Altarelli admits. His group is developing new detectors, but, he says, "I'm not saying 'Ah, we'll be ready anytime.' " But with a few years' leeway, they hope to be ready to catch up to the results likely to be coming out of Menlo Park and Hyogo Prefecture.

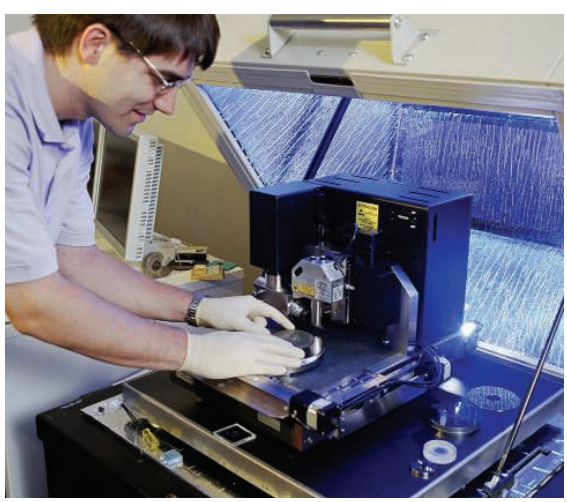
–DENNIS NORMILE

U.S. RESEARCH POLICY

Congress Cancels Contentious Program to Bolster Industry

The Advanced Technology Program met its goals, argue its supporters—and critics say that's why it needed to be killed

Five years ago, Peter Fiske was running out of time and money. The physicist-turned-entrepreneur and a partner had set up a company, called RAPT Industries, to commercialize a new technique to etch semiconductors. But venture capitalists weren't interested. They said the technology, based on research by Fiske's partner at Lawrence Livermore National Laboratory in California, wasn't mature. And family loans went only so far. So Fiske* and his partner turned to Uncle Sam.



A helping hand. Jude Kelley measures precision optical surfaces at RAPT Industries, which has benefited from ATP funding.

CREDIT: RAPT INDUSTRIES INC.

They applied to the Advanced Technology Program (ATP)—a U.S. Commerce Department program that helps companies develop promising but risky technologies—and won a $2 million grant for research and testing. The money bridged the so-called valley of death between the lab bench and a marketable product—one of the main goals of ATP—and last year, RAPT Industries recorded $700,000 in sales. But Fiske's company may be among the last to benefit from ATP: After 16 years and more than $2 billion in tax money, the program is closing up shop.

Good riddance, say its critics, who believe that market forces, not a government agency, should determine the commercial fate of new technologies. That view holds sway in the House of Representatives, which has voted eight times to kill the program. But supporters, including a handful of senators who have succeeded until this year in rescuing the program, have argued that ATP is needed to ensure that promising technologies such as Fiske's don't die on the vine. And a drumbeat of studies conducted throughout ATP's life attest to its effectiveness.

For a program that has so angered free-market advocates, ATP has a surprising pedigree: It was launched during the Reagan Administration, and President George H. W. Bush provided the initial funding. Its biggest backer was Senator Fritz Hollings (D–SC), who in the late 1980s saw government subsidies as the primary driver behind Japan's ascendancy in the field of computer chips. Hollings proposed a federal initiative to make U.S. industry more competitive, based at the well-respected National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland, so that its scientists could offer the best technical peer review of what companies wanted to do.

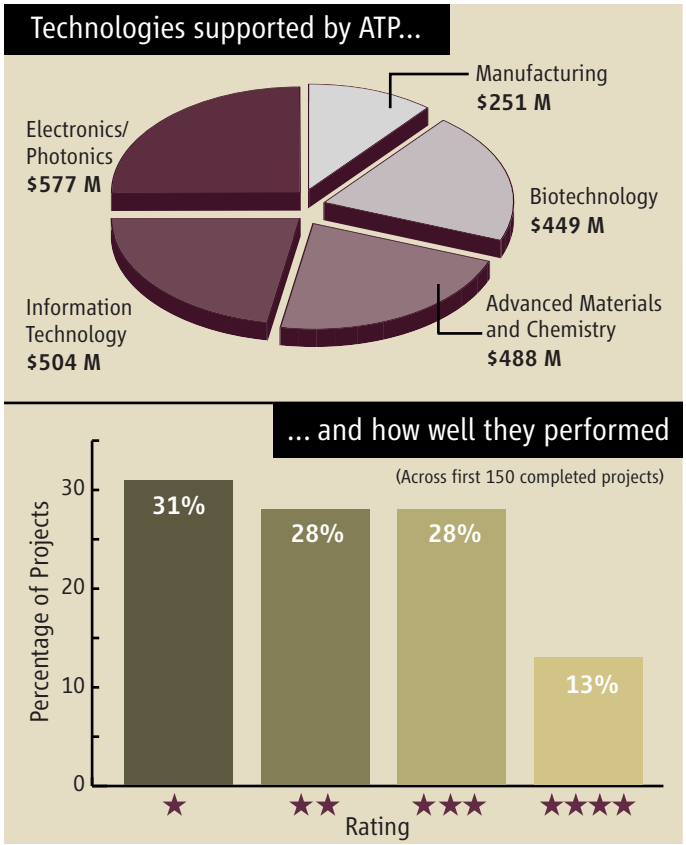
* Fiske also writes a column for *Science*'s careers Web site, ScienceCareers.org.

Companies were required to finance nearly one-half the research costs of projects, which would be continually assessed by NIST staffers to make sure they remained on target.

"Hollings was absolutely insistent that this be a merit-reviewed, nonporked program," says Pat Windham, a former aide to the now-retired senator, referring to the popular legislative practice of designating money for projects—from roads to research—that have not been vetted by the agency that will run them. "What we were looking for were [serious] proposals from industry," says Windham, now a policy consultant. "The [idea] was that these things were sufficiently long-term that companies wouldn't have funded the work on their own."

Few question that ATP's track record of 768 funded projects contains some real winners. One home run involved Affymetrix, a Santa Clara, California, biotech company founded in 1993 to sell chips to carry out automated genomic analyses. The $31 million in ATP funds the company and its partners received between 1995 and 2000 helped Affymetrix hire academic scientists and develop new software and equipment. The funds "validated our technology and helped accelerate the development of [gene chips]," says Affymetrix official Robert Lipshutz. Last year, the gene-array company racked up $367 million in sales.

The benefits of ATP projects have extended far beyond the companies themselves. Several studies examining a total of 14 projects have claimed an economic return that exceeded $1.2 billion for the $87 million spent by the government. In 1998, a study by the Research Triangle Institute in Research Triangle Park, North Carolina, estimated that seven successful tissue-engineering projects that received roughly $15 million from ATP saved society $34 billion in reduced morbidity and lower medical costs.



Keeping score. The Advanced Technology Program has spent $2.3 billion to support 768 projects, with varying degrees of commercial success.

CREDIT: ADAPTED FROM ATP 2006 REPORT ON ECONOMIC PROGRESS, NIST

Ideological battles

But despite such success stories, conservatives remain convinced that the government has no business subsidizing commercial research and development. "Companies should have every incentive to fund this kind of profitable research on their own," wrote Brian Riedl of the libertarian Heritage Foundation this year, summarizing what ATP opponents have long characterized as "corporate welfare." Others said the program amounted to picking winners, a task for which the government was ill-suited.

The debate first came to a boil in 1995, when former president Bill Clinton tried to increase the program's budget sixfold as part of his Administration's efforts to make U.S. industry more competitive globally. Instead, the new Republican majority in Congress voted to kill the program outright, and only a presidential veto saved it. Robert Walker, then chair of the House Science Committee and a close ally of speaker Newt Gingrich, says Republican leaders "had a hard time justifying ATP as real science" because "the program was designed to bring [existing] technology to the market."

He and other critics felt it also tilted the playing field. A 2000 General Accounting Office report, for example, found that in the case of one $2 million award for a tissue-engineering project, "many competitors were attempting to achieve similar broad research goals." The ATP grant made the program "not fair to taxpayers or competitors," says Riedl.

Not so, says Robert Boege of the Alliance for Science & Technology Research in America in Washington, D.C. There's "a long and venerable tradition" of government sponsorship of emerging technologies, says Boege, citing the musket, telegraph, and railroad industries. "The Internet itself is a result of public science," he adds.

Economist Adam Jaffe of Brandeis University in Waltham, Massachusetts, says that government-funded corporate research will inevitably overlap with some private activities. "ATP at least made an effort" to prevent that from happening through its rigorous peer review, says Jaffe. The program also tried to be transparent, assigning projects from zero to four stars based on the industrial progress or technical innovations they came up with (see graph).

But that openness also left it more vulnerable to critics. "Even those who think the government has a role in funding [corporate] R&D should be concerned that two-thirds of the programs have no return," says Riedl about an ATP evaluation of its first 150 completed projects.

Supporters argued to little avail that some failures are inevitable and that venture capital companies would celebrate such a 2:1 ratio of failures to successes. "They're apparently the only program in the federal government that has that problem," NRC staffer Charles Wessner says dryly, wondering why ATP was singled out for such criticism.

Hollings's retirement in 2004 put ATP in jeopardy, and the fact that business lobbyists never fought hard to preserve it sealed its fate. "We all knew that when Hollings left it was going to be bye-bye time for ATP," says former House Science Committee staffer Olwen Huxley.

Supporters believe strongly that the concept remains valid, however. Technology bureaucrats in Finland and Sweden, among other European nations, have expressed interest in some of ATP's funding and evaluation techniques, and true believers fantasize about a U.S. revival should the Democrats win control of one or both houses of Congress after next week's midterm elections. In the meantime, ATP's rise and fall shows both the political allure of taxpayer-funded corporate research and the difficulty of keeping the dollars flowing.

–ELI KINTISCH

LETTERS

edited by Etta Kavanagh

How Protected Are Coral Reefs?

THE POLICY FORUM "CORAL REEFS AND THE GLOBAL NETWORK OF Marine Protected Areas" (C. Mora *et al.*, 23 June, p. 1750) draws attention to the vulnerability of coral reef Marine Protected Areas (MPAs) to human activities. The authors evaluated the exposure of coral reef reserves to poaching and to external threats (pollution, erosion, overexploitation, and shoreline development) using a risk index. Remarkably, neither the authors nor the source of their risk index (*1*) identify biological invasion (the introduction of nonindigenous organisms) as a significant threat. I believe this reflects the conventional wisdom that tropical regions and, in particular, highly diverse systems like coral reefs are largely immune to invasion.

However, the few studies that have investigated nonindigenous species on coral reefs found that, although they comprise a minor proportion of the total diversity, invaders are capable of damaging reef ecosystems. Severe impacts of invasive algae and pathogens have been documented (*2*, *3*), and cases involving other organisms continue to accrue. In recent years, an octocoral from the western Atlantic and a sponge from Indonesia have been overgrowing and killing native corals in Hawaii (*4*). Similarly, a stony coral from the Indo-Pacific has begun to foul reefs off Florida and Brazil (*5*).

The magnitude of the problem is certainly underestimated, as the origins of large numbers of invertebrates, bacteria, and viruses occurring on reefs are unknown. Furthermore, several thousand species are being moved across the world in ballast tanks and on the hulls of ships (*6*), and aquarium releases are contributing to the spread of species in tropical regions (*7*). Hence, the threat posed by biological invasion is unlikely to diminish and should therefore be considered in analyses of the effectiveness of MPAs.

ANTHONY RICCIARDI

Redpath Museum, McGill University, Montreal, QC H3A 2K6, Canada.

References

1. D. Bryant, L. Burke, J. McManus, M. Spalding, *Reefs at Risk: A Map-Based Indicator of Potential Threats to the World's Coral Reefs* (World Resources Institute, Washington, DC, 1998).
2. J. E. Smith, *Ecol. Lett.* **9**, 835 (2006).
3. C. D. Harvell *et al.*, *Science* **285**, 1505 (1999).
4. S. E. Kahng, R. W. Grigg, *Coral Reefs* **24**, 556 (2005).
5. D. Fenner, K. Banks, *Coral Reefs* **23**, 505 (2004).
6. J. T. Carlton, in *Invasive Species and Biodiversity Management*, O. T. Sandlund *et al.*, Eds. (Kluwer Academic Publishers, Dordrecht, Netherlands, 1999), pp. 195–212.
7. H. S. Meister *et al.*, *Southeast. Nat.* **4**, 193 (2005).

THE POLICY FORUM "CORAL REEFS AND THE global network of Marine Protected Areas" by C. Mora *et al.* (23 June, p. 1750) underestimates the complexity of the conservation challenge.

First, the analysis does not factor in the impacts of some of the most important pervasive global anthropogenic stressors on coral (*1*) that penetrate Marine Protected Area (MPA) boundaries via terrestrial, atmospheric, and oceanic avenues (*2*). These include increasing sea surface temperatures and associated coral bleaching, contagious coral disease, and potential ocean acidification (*3*).

Second, although Mora *et al.* recognize the inadequacies of management and enforcement within MPAs themselves, they do not integrate the potential impacts of larger, and equally important, political, economic, and sociological forces into their analysis. For example, it is possible to establish a perfect global MPA network using all the best science, but still fail to protect coral reefs if you do not have high and sustained political and community capacity at local and national levels (*2*). Special interest groups that make campaign contributions and gain favorable permit decisions from politicians (low political capacity) can ruin the best scientifically designed MPA network in a short period of time. Likewise, if local residents do not have a conservation ethic (low community capacity), no amount of regulation and enforcement will protect coral reef resources in the long run from stressors like poaching. Low political and community capacity situations are more the rule than the exception in the MPA world.

We all have a vested interest in making MPAs effective tools for conserving coral, enhancing fisheries, and conserving related reef biodiversity, but to make the MPA tool effective for conserving coral, we must reduce the root causes of pervasive global anthropogenic stressors (*4*). This starts with changing our own personal behavior and extends to making larger political, cultural, and economic improvements. These include, but are not limited to, citizens demanding governmental enforcement of existing environmental regulations, voters participating in the political process, and stockholders demanding environmentally responsible business behavior. None of these tasks are easy or ever complete.

Any reassessment of global-scale conservation strategies for coral reefs, in this era of global economies, climate change, and interconnected ecosystems, must focus on reducing the root cause of stressors on coral and on improving political and community capacity, because the effectiveness of any global MPA network is inextricably linked to success in these critical areas.

What the analysis of Mora *et al.* does show clearly is that the use of the term Marine "Protected" Area is truly a misnomer. The term Marine "Managed" Area is more appropriate to describe this conservation tool. The MPA term should only be used if real "protection" can be biologically certified over time (*2*).

STEPHEN C. JAMESON

Chairman and President, Coral Seas Inc.–Integrated Coastal Zone Management, 4254 Hungry Run Road, The Plains, VA 20198, USA. E-mail: sjameson@coralseas.com

References

1. D. Bryant, L. Burke, J. McManus, M. Spalding, *Reefs at Risk: A Map-Based Indicator of Potential Threats to the World's Coral Reefs* (World Resources Institute, Washington, DC, 1998), p. 17.
2. S. C. Jameson, M. H. Tupper, J. M. Ridley, *Mar. Pollut. Bull.* **44** (no. 11), 1177 (2002) (see www.coralseas.com/press.html).
3. S. C. Jameson, J. W. McManus, M. D. Spalding, *State of the Reefs: Regional and Global Perspectives* (International Coral Reef Initiative, U.S. Department of State, Washington, DC, 1995) (see www.coralseas.com/press.html).
4. P. M. Scanlan, *The Dolphins Are Back: A Successful Quality Model for Healing the Environment* (Productivity Press, Portland, OR, 1998).

IN THEIR POLICY FORUM "CORAL REEFS AND the global network of Marine Protected Areas" (23 June, p. 1750), C. Mora *et al.* discuss the destruction of coral reefs and international agreements to protect these fragile ecosystems. The authors based their analysis on the 2005 version of the World Database on Protected Areas (WDPA) (*1*). This database is maintained by the United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC) (*2*) in collaboration with the World Conservation Union on behalf of a consortium of organizations. We suggest that their analysis could have been substantially improved if they had used the more recent WDPA data that are available online from our collaboration with the University of British Columbia Sea Around Us Project (*3*).

The WDPA is a primary source of protected areas information for many research activities. It serves a wide range of stakeholders, including governmental and intergovernmental bodies, policy advisors, researchers, managers, and private-sector decision-makers. The WDPA is compiled from protected areas information provided by competent agencies, with additional input from researchers and professional experts in the field. It now contains standardized data for 233 countries and territories, including marine and coastal areas, and is freely available for noncommercial purposes in keeping with the principles of the Conservation Commons (*4*). Although annual updates are released on CD for distribution at relevant international fora, users can access the most up-to-date information online. Given the complexity of the WDPA content, we encourage users to seek our advice directly to ensure that they are using the most recent data sets and that it is interpreted appropriately. In return, we welcome access to any relevant new data that researchers can provide so that we can improve the resource for other researchers and decision-makers.

ED MCMANUS, CHARLES BESANÇON, TIM JOHNSON

United Nations Environment Programme, World Conservation Monitoring Centre, Cambridge, CB3 0DL, UK.

References

1. See www.unep-wcmc.org/wdpa.
2. See www.unep-wcmc.org.
3. See www.seaaroundus.org.
4. See www.conservationcommons.org.

IN "CORAL REEFS AND THE GLOBAL NETWORK of Marine Protected Areas" (Policy Forum, 23 June, p. 1750), C. Mora *et al.* suggest that only 2% of the world's coral reefs are adequately protected. We believe that the authors have set impossibly high standards for "adequacy" and have misdirected attention from the real problems facing coral reefs and the even greater needs for marine protection of other habitats. For example, Australia's Great Barrier Reef is described as "partially protected." The only recent published global reef map (*1*) suggests that this reef represents almost 14% of the world's coral reefs, and over one-third of it has been designated as strictly protected. Although this reef is still subject to pressures from climate change and runoff from the mainland, a simple classification of this flagship MPA as inadequate seems to be making a statement to policy-makers that they can never succeed.

Coral reefs are, in fact, the best protected of all marine and coastal habitats. Using the World Database on Protected Areas (*2*), together with recent updates, we estimate that approximately 22.6% of all reefs fall within some classification of legal protection, while 11.4% fall within classes of stricter management regimes (IUCN management categories I to IV). These are crude measures and the effectiveness of many sites may be called into question, but we cannot doubt that considerable progress has been made. In fact, there has been a high positive selection for reef areas—overall, only 4.3% of shelf areas (above 200 m) fall within some level of protection and only 1.9% within stricter levels of protection. Other critical marine habitats—such as kelp forests, deep coral communities, seagrasses, seamounts, and the vast expanses of the high seas—are far less protected.

We remain far from the goal of achieving representative networks of MPAs by 2012 (*3*), even for coral reefs, but attention also needs to be focused more broadly than simple coverage statistics. We should be trying to design MPA networks that are resilient to the many ex situ influences that do not respect liquid boundaries in the ocean (pollution, disease, overharvesting of entire fish stocks, and the many influences of climate change). Such MPA networks further need to be placed into a more integrated framework for management, covering, inter alia, watershed-based management, ecosystem-based management of fisheries, and globally targeted policy changes in carbon emissions.

MARK SPALDING,[1] GRAEME KELLEHER,[2] TIMOTHY BOUCHER,[1] LUCY FISH[3]

[1]Global Conservation Approach Team, The Nature Conservancy, Arlington, VA 22203–1606, USA. [2]Chair, World Commission on Protected Areas High Seas MPA Task Force, Canberra ACT 2614, Australia, and former Chair, Great Barrier Reef Marine Park Authority. [3]United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC), Cambridge CB3 0DL, UK.

References

1. M. D. Spalding, C. Ravilious, E. P. Green, *World Atlas of Coral Reefs* (Univ. of California Press, Berkeley, CA, 2001).
2. WDPA custodian: UNEP World Conservation Monitoring Centre, see also www.unep-wcmc.org/wdpa/.
3. WSSD (United Nations, 2002).

Response

WE PROVIDED A GLOBAL AUDIT OF THE MANagement effectiveness of Marine Protected Areas (MPAs) containing coral reefs. We found that less than 0.01% of the world's coral reefs are within MPAs that fully protect reef diversity from threats due to poaching, overfishing, coastal development, and pollution. Ricciardi and Jameson suggest that we overestimated the protection received by coral reefs because invasive species and climate change were not considered. We agree that those two threats are likely adding to the worldwide vulnerability of coral reefs. However, our paper was not intended to quantify this vulnerability, but to assess the effectiveness of MPA management.

Many threats to coral reefs are local (e.g., overfishing, pollution, and coastal development) and can be policed as part of the management plan of MPAs. However, there are other threats (e.g., climate change and invasive species) that are not local and thus are more difficult to police or even monitor from a MPA. Controlling the effects of climate change and invasive species is unlikely to be an effective function of MPAs, but if we consider them as such, that will only worsen the current management situation of MPAs worldwide. Jameson further suggests that our analysis failed to consider political, economical, and sociological data, which do influence MPA effectiveness. We agree. However, these factors are likely to be reflected in the levels of

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

poaching inside MPAs, which we did quantify.

MPAs have proven effective at reducing the effects of fishing. However, they have to be complemented with additional approaches to reduce other human pressures. Increasing general public awareness of environmental problems is important. It is likely that if we overcome the indifference of governments and the general public to environmental issues (*1*), we could reduce the impact of human stressors and achieve a broader protection of biodiversity.

McManus and colleagues suggest that our analysis "could have been substantially improved" if we had used their most up-to-date version of the World Database on Protected Areas (WDPA). We recognize the great value of this database, acknowledge that it was important for our analysis, and were aware of the current attempt to verify and update it. However, we realized that such a new database was not going to be ready in time for our analysis and therefore decided to carry out an independent review of the database. For this process of verification, we used recent reports and contacted over 1000 researchers and managers in 103 countries. Our analysis included these corrections (verification and updating) to the 2005 WDPA relevant to coral reefs, and therefore we doubt that waiting for a new and better database would have "substantially" improved our results.

In our analysis, each MPA was classified in one of four categories of effectiveness ranging from adequate to very limited conservation status. The category defined as "adequate" included MPAs that were mostly no-take with no or low levels of poaching and low to medium risk and were variable in size and isolation. We found that only 2% of the world's coral reefs are within MPAs categorized as adequate. Spalding and colleagues claim that this category includes quality standards that are impossible to achieve and therefore our results are a message to policy-makers that they can never succeed. We disagree. First, compelling evidence suggests that MPAs have to be no-take and have to be minimally affected by external risk to provide appropriate protection to coral reefs (*2*, *3*). So we consider that the attributes we define as "adequate" should be the minimum characteristics that an MPA should have to be effective. Second, we do not believe that the standards we set as adequate are impossible to reach. The recent upgrade to no-take status of the Northern Hawaiian Islands is a good indicator that setting aside large areas from the effects of fishing is possible. Reductions in the impact of external risks such as runoff are also achievable, and advances are being made in large areas like the Florida Keys and the Great Barrier Reef.

Spalding and colleagues claim that our study "misdirected attention from the real problems facing coral reefs and the even greater needs for marine protection of other habitats." We did not make any claim about the status of MPA effectiveness in other marine habitats. It is very likely that the situation we described for coral reef MPAs is occurring in other habitats, but what that suggests is the great need for effective conservation of all marine habitats. MPAs are one of the main approaches used for the conservation of coral reefs worldwide, and our paper "directs" attention to the problems they have in achieving effective protection. That is not to say that MPAs alone are going to prevent the large plethora of threats affecting coral reefs and that other approaches should not be used.

Finally, Spalding and colleagues argue that coral reefs are the best protected of all

marine and coastal habitats. One of the main messages of our paper is the need to differentiate between quantity and quality of protection by MPAs. Establishing parks on paper can easily increase the quantity of protection, but that coverage is not effective and may provide a false sense of security. Our study shows that this is the case for coral reefs. Therefore, the statement by Spalding and colleagues that "[c]oral reefs are, in fact, the best protected of all marine and coastal habitats" should be taken with care, because although 18.7% of the world's coral reefs are within MPAs, only 2% are adequately protected. This suggests that MPAs worldwide are, for the most part, poorly effective and that current efforts to reverse the existing crisis of coral reefs fall far short of what is required to save these most diverse of all marine habitats.

CAMILO MORA,[1,2]* SERGE ANDRÉFOUËT,[3] MARK J. COSTELLO,[2] CHRISTINE KRANENBURG,[4] AUDREY ROLLO,[2] JOHN VERON,[5] KEVIN J. GASTON,[6] RANSOM A. MYERS[1]

[1]Department of Biology, Dalhousie University, Halifax, NS B3H 4J1, Canada. [2]Leigh Marine Laboratory, University of Auckland, Post Office Box 349, Warkworth, New Zealand. [3]Institut de Recherche pour le Développement, Boite postale A5-98848, Noumea cedex, New Caledonia. [4]Institute for Marine Remote Sensing, University of South Florida, St. Petersburg, FL 33701, USA. [5]Australian Institute of Marine Sciences, Townsville 4810, Australia. [6]Biodiversity and Macroecology Group, Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK.

*Author for correspondence. E-mail: moracamilo@hotmail.com

References

1. J. R. Miller, *Trends Ecol. Evol.* **20**, 430 (2005).
2. P. F. Sale *et al.*, *Trends Ecol. Evol.* **20**, 74 (2005).
3. D. M. Stoms *et al.*, *Front. Ecol. Environ*. **3**, 429 (2005).

CORRECTIONS AND CLARIFICATIONS

News of the Week: "On your mark. Get set. Sequence!" by E. Pennisi (13 Oct., p. 232). Ewan Birney is not at the Wellcome Trust Sanger Institute as stated, but at the European Bioinformatics Institute, which is a part of the European Molecular Biology Laboratory.

News of the Week: "Perelman declines math's top prize; three others honored in Madrid" by D. Mackenzie (25 Aug., p. 1027). The article identified Richard Hamilton's affiliation incorrectly as the State University of New York at Stony Brook; he is at Columbia University. Also, the manuscript by Bruce Kleiner and John Lott appeared May 25, not in June, and the manuscript by Huai-Dong Cao and Xi-Ping Zhu was dated June 2006, not April.

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Rapid Uplift of the Altiplano Revealed Through ^{13}C-^{18}O Bonds in Paleosol Carbonates"

T. Sempere, A. Hartley, P. Roperch

Based on stable isotope measurements, Ghosh *et al*. (Reports, 27 January 2006, p. 511) concluded that the Bolivian Altiplano uplifted 3 to 4 kilometers between ~10.3 and ~6.7 million years ago as a result of gravitational loss of dense lithosphere. This result stands at odds with current geological knowledge of the Central Andes, and we propose a test for the reliability of the paleoaltimetry method.

Full text at www.sciencemag.org/cgi/content/full/314/5800/760b

RESPONSE TO COMMENT ON "Rapid Uplift of the Altiplano Revealed Through ^{13}C-^{18}O Bonds in Paleosol Carbonates"

John Eiler, Carmala Garzione, Prosenjit Ghosh

Clumped-isotope thermometry measurements of carbonate samples deposited in the Bolivian Altiplano as early as 28.5 million years ago and buried up to ~5000 meters deep exhibit no relationship between burial depth and apparent temperature, and largely yield temperatures within error of plausible Earth-surface conditions. These results counter the predictions of Sempere *et al*. and support our previous conclusions regarding the uplift of the Altiplano.

Full text at www.sciencemag.org/cgi/content/full/314/5800/760c

Comment on "Rapid Uplift of the Altiplano Revealed Through ^{13}C-^{18}O Bonds in Paleosol Carbonates"

T. Sempere,[1]* A. Hartley,[2] P. Roperch[3]

Based on stable isotope measurements, Ghosh *et al*. (Reports, 27 January 2006, p. 511) concluded that the Bolivian Altiplano uplifted 3 to 4 kilometers between ~10.3 and ~6.7 million years ago as a result of gravitational loss of dense lithosphere. This result stands at odds with current geological knowledge of the Central Andes, and we propose a test for the reliability of the paleoaltimetry method.

Ghosh *et al*. (*1*) reconstructed the elevation history of the Altiplano plateau in the Bolivian Andes using a thermometer based on the temperature-dependent binding rate of ^{13}C and ^{18}O isotopes in carbonate minerals. Their measurements indicate that the Altiplano lay between –400 and 0 m from 11.4 to 10.3 million years ago (Ma) and rose to its current altitude at an average rate of 1.03 ± 0.12 mm per year between ~10.3 and ~6.7 Ma. Ghosh *et al*. concluded that such a rapid uplift was likely to have been produced by gravitational loss of dense lithosphere to the asthenosphere (delamination), but this scenario disagrees with current geological knowledge of the Central Andes.

The idea that part of the Bolivian Altiplano was at or below sea level as late as ~11 Ma disagrees with the common view that Andean orogeny started in western Bolivia either ~26 Ma (*2*) or ~40 Ma (*3*, *4*), and with geomorphic evidence that the volcanic highlands west of the Altiplano were above 2000 m as early as 20 to 17 Ma (*5*). Furthermore, forearc strata document that these highlands underwent uplift between ~40 and 10 Ma (*6*). The fault-bounded Corque Basin displays high compacted sedimentation rates [970 m per million years (My) between 12 and 9 Ma (*7*), decreasing to 337 m/My between 9 and 6 Ma (*8*)] and can be seen as a pull-apart basin (*9*) whose surface was at a substantially lower altitude than surrounding highlands. Therefore, generalization of paleoaltitudes reconstructed in the basin (*1*) to the entire Altiplano may be inappropriate.

Ghosh *et al*. (*1*) argued that their proposed uplift history is consistent with paleobotanical evidence (*10*). However, the current paleoaltimetry method based on fossil leaf morphology systematically underestimates high altitudes (*11*). Therefore, Low Miocene paleoaltitudes reported for Andean Bolivia using this method (*10*) may also be underestimations and cannot be invoked to support the results in (*1*).

Crustal thickening in the Central Andes is widely believed to have been caused by tectonic shortening (*12*). On the contrary, Ghosh *et al*. (*1*) contend that this process is too slow to account for the rapid uplift of the Altiplano implied by their results. Instead the authors suggest crustal delamination, removal of dense lower crust and/or mantle lithosphere, as a more plausible mechanism. However, this process can only occur when the lower part of the lithosphere has become gravitationally unstable as a result of thickening (*13*). Delamination below the Altiplano (*1*) should thus have been a consequence of thickening. However, if the Corque Basin was indeed at or below sea level at ~11 Ma (*1*), the crust—which is now ~55 km thick (*14*)—had not been thickened by then. Because an unthickened crust implies an unthickened lithosphere, it is difficult to explain why the lower lithosphere would start to delaminate before thickening. Even if thickening of the Altiplano crust started at 10.3 Ma with simultaneous "slow" delamination, it is unclear what process triggered thickening at that time.

Can soil paleotemperatures, and hence paleoaltitudes, be securely deduced from isotope-geochemical measurements? Ghosh *et al*. (*1*) assumed that the carbonate nodules they analyzed were devoid of diagenetic signal, yet they reported one sample (04BL69) from the 10.3 to 11.4 Ma interval that yielded an apparent paleotemperature of 50.3° ± 4.9°C and acknowledged that this was likely due to cryptic recrystallization during burial. The samples from this interval were subject to minimum burial depths of between 2200 and 3400 m (*8*) and thus to temperatures of 60° to 90°C (adopting a conservative estimate of 30°C/km for the geothermal gradient). We believe it unlikely that only one sample was selectively affected by burial metamorphism and that samples above and below were not.

We propose a simple test to determine whether a burial heating component is indeed present in the geochemical signal. The ~11.4 to 5.8 Ma, ~3.5-km-thick section analyzed by Ghosh *et al*. (*1*) [and (*8*)] is only the uppermost part of the ≥12-km-thick (*15*), 55.5-Ma continental succession that crops out in the Corque syncline. A ~4.7-km-thick part of this succession, partly overlapping with the former (*1*, *8*), was reliably dated 14.5 to 9.0 Ma by magnetostratigraphy (*7*) and displays facies, including carbonate nodules, somewhat similar to the ~11.4 to 5.8 Ma succession. Collecting samples down-section and processing them by the method used by Ghosh *et al*. (*1*) would show whether apparent paleotemperatures keep growing down-section or not, and thus refute or validate their method.

Because isotopic resetting may occur during burial diagenesis of paleosol nodules, the geochemical methods used by Ghosh *et al*. (*1*) should have been robustly validated by thorough down-section sampling before drawing conclusions about the history and mechanisms of uplift in the Central Andes. If burial is proved to have modified the geochemical signal as we predict, the reported paleoaltitude estimates from at least the 11.4 to 10.3 Ma interval will need to be reevaluated and the rapid and late Andean uplift proposed by Ghosh *et al*. reconsidered.

[1]Institut de Recherche pour le Développement and Laboratoire "Mécanismes et Transferts en Géologie," Observatoire Midi-Pyrénées, Université Paul Sabatier, 31400 Toulouse, France. [2]School of Geosciences, University of Aberdeen, Aberdeen AB24 3UE, UK. [3]Institut de Recherche pour le Développement and Géosciences Rennes, Université de Rennes, 35042 Rennes, France.

*To whom correspondence should be addressed. E-mail: sempere@lmtg.obs-mip.fr

References

1. P. Ghosh, C. N. Garzione, J. M. Eiler, *Science* **311**, 511 (2006).
2. T. Sempere, G. Hérail, J. Oller, M. G. Bonhomme, *Geology* **18**, 946 (1990).
3. S. Lamb, P. Davis, *Nature* **425**, 792 (2003).
4. B. K. Horton, *Tectonics* **24**, TC3011 10.1029/2003TC001619 (2005).
5. M. Sébrier, A. Lavenu, M. Fornari, J.-P. Soulas, *Géodynamique* **3**, 85 (1988).
6. M. Farías, R. Charrier, D. Comte, J. Martinod, G. Hérail, *Tectonics* **24**, TC4001 (2005).
7. P. Roperch, G. Hérail, M. Fornari, *J. Geophys. Res.* **104**, 20415 (1999).
8. C. N. Garzione, P. Molnar, J. C. Libarkin, B. J. MacFadden, *Earth Planet. Sci. Lett.* **241**, 543 (2006).
9. S. Rousse, S. Gilder, M. Fornari, T. Sempere, *Tectonics* **24**, 10.1029/2004TC001760 (2005).
10. K. M. Gregory-Wodzicki, *Geol. Soc. Am. Bull.* **112**, 1091 (2000).
11. E. A. Kowalski, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **188**, 141 (2002).
12. B. L. Isacks, *J. Geophys. Res.* **93**, 3211 (1988).
13. R. W. Kay, S. M. Kay, *Tectonophysics* **219**, 177 (1993).
14. S. L. Beck *et al.*, *Geology* **24**, 407 (1996).
15. P. Rochat, G. Hérail, P. Baby, G. Mascle, O. Araníbar, *C. R. Acad. Sci. Paris* **327**, 769 (1998).

21 July 2006; accepted 4 October 2006
10.1126/science.1132837

Response to Comment on "Rapid Uplift of the Altiplano Revealed Through ^{13}C-^{18}O Bonds in Paleosol Carbonates"

John Eiler,[1]* Carmala Garzione,[2] Prosenjit Ghosh[1]

Clumped-isotope thermometry measurements of carbonate samples deposited in the Bolivian Altiplano as early as 28.5 million years ago and buried up to ~5000 meters deep exhibit no relationship between burial depth and apparent temperature, and largely yield temperatures within error of plausible Earth-surface conditions. These results counter the predictions of Sempere *et al*. and support our previous conclusions regarding the uplift of the Altiplano.

Sempere *et al*. (*1*) suggest that the temperatures recorded by carbonate clumped-isotope thermometry in 11.4 to 10.3 million-year-old soil nodules from the northern Altiplano (*2*) reflect partial resetting during burial rather than deposition at low altitude. Their arguments include a testable prediction: If the soil carbonates in question underwent partial resetting during burial, then more deeply buried samples from the same or related sections should be even more strongly reset, yielding apparent temperatures above any plausible depositional temperature.

Figure 1 presents the results of carbonate clumped-isotope thermometry analyses for 32 soil and lacustrine carbonates from the northern Altiplano. These data include those in (*2*) as well as new measurements that are part of a broader ongoing study of modern and ancient carbonates [generated using the same analytical methods described in (*2*)]. This expanded suite includes soil carbonates deposited between 28.5 and 0 million years ago (Ma) and buried between 0 and ~5000 m deep, as well as lake carbonates of similar age and burial depth. Age estimates for the new measurements are based on recently published magnetostratigraphy (*3*) and previously published $^{40}Ar/^{39}Ar$ dates (*4*) of tuffs within the Corque and Tambo Tambillo sections, and on magnetostratigraphy (*5*) within the Salla section. We estimated maximum burial depths for each sample based on our own measured sections near Callapa (*3*) and estimated section thicknesses in the Tambo Tambillo and Salla areas (*4*, *6*).

The data presented in Fig. 1 exhibit no systematic relationship between apparent growth temperature and burial depth, are generally within analytical uncertainty of earth-surface temperatures [the only noteworthy exception was reported and discussed in (*2*)], and include relatively low temperatures in samples far older and more deeply buried than those reported by (*2*)—i.e., the temperatures of 16.9°C and 21.5°C found in 23.6- to 23.7-Ma soil carbonates that were buried to ~5000 m. Moreover, we observe no systematic difference between surface-deposited carbonates of different types (i.e., soil versus lacustrine). Variations in temperature within this suite stem from a variety of factors, including primary differences in paleoaltitude and paleoclimate [discussed for many of the Callapa samples in (*2*)], unusual diagenetic resetting [e.g., the one high-temperature Callapa sample discussed in (*2*)], and analytical uncertainties. It is beyond the scope of this reply to discuss all of these issues in detail. Nevertheless, these data contradict the predictions of Sempere *et al*. and, more generally, lend no support to the suggestion that burial metamorphism has systematically reset the growth temperatures of Altiplano soil carbonates. For this reason, we maintain that the difference in average apparent temperature between 11.4 and 10.3 Ma and post-6.7-Ma soil carbonate suites reported in (*2*) reflects a difference in their temperatures of deposition and thus constrains paleoaltitudes using methods (and with uncertainties) that have already been discussed (*2*).

Sempere *et al*.'s geomorphic and stratigraphic arguments against a late Miocene date for uplift of the northern Altiplano are relevant but contain no quantitative paleoaltitude determinations and say nothing specific about the mid- to late-Miocene paleoaltitude of the Altiplano. Sempere *et al*. recognize that the Western Cordillera could have extended to higher altitudes than the Altiplano (as they do today); we suggest it is also possible that mid-Miocene altitudes in any or all of these regions might have been higher or lower than Oligocene and early Miocene altitudes. It will be difficult to know how to evaluate these issues until there is a quantitative database documenting temporal and spatial variations of paleoaltitudes across the Andean orogen.

Sempere *et al*.'s critique of paleoaltimetry based on fossil leaf assemblages has no direct

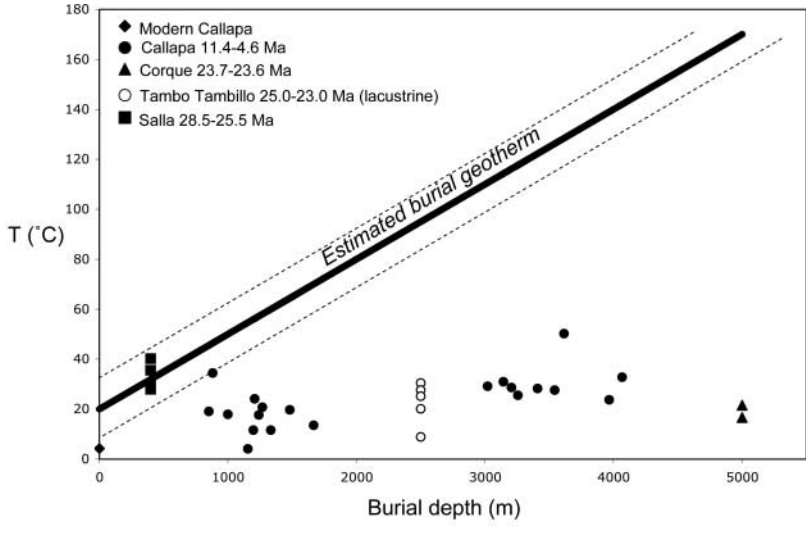


Fig. 1. Apparent growth temperatures for various Altiplano carbonates based on clumped-isotope thermometry plotted as a function of estimated maximum burial depth. Symbols discriminate among soil carbonates from sections near Callapa, Corque, and Salla and lacustrine carbonates from near Tambo Tambillo, as indicated by the legend. The heavy solid line indicates an estimated burial geotherm, assuming a surface temperature of 20°C and a gradient of 30°C per km. The dashed lines define a ±10°C offset from this trend, which we consider a reasonable estimate of its uncertainty. Carbonates deposited on or near the surface of the Altiplano within the past 28.5 million years and buried to 5000 m or less exhibit no systematic relationship between apparent temperature and burial depth and show no evidence for pervasive resetting of deeply buried samples.

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [2]Department of Earth and Environmental Sciences, University of Rochester, Rochester, NY 14627, USA.

*To whom correspondence should be addressed. E-mail: eiler@gps.caltech.edu

bearing on the Ghosh *et al*. study (*2*). Although we noted that clumped-isotope thermometry results broadly agree with paleobotanical altimetry, our central arguments do not depend on this issue.

We do not agree with Sempere *et al*. (*1*) that removal of mantle lithosphere requires previous crustal thickening beneath the Altiplano. The Eastern Cordillera preserves the largest documented shortening in the Andes (*7*, *8*) and is the most plausible candidate for the locus of development of an unstable lower-crustal and/or lithospheric-mantle root. Gravitational removal of this material could have led to simultaneous surface uplift of the eastern Altiplano and Eastern Cordillera and lower crustal flow from the Eastern Cordillera to the Altiplano, thickening the crust beneath the Altiplano. This scenario is only one of several that cannot be discounted using existing constraints. Nevertheless, it is an example of a process that is consistent with both the paleoaltitude reconstructions of (*2*) and the physics that govern convective removal of lithosphere, crustal thickness, and isostasy.

References and Notes

1. T. Sempere, A. Hartley, P. Roperch, *Science* **314**, 760 (2006); www.sciencemag.org/cgi/content/full/314/5800/760b.
2. P. Ghosh, C. N. Garzione, J. Eiler, *Science* **311**, 511 (2006).
3. C. N. Garzione, P. Molnar, J. C. Libarkin, B. J. MacFadden, *Earth Planet. Sci. Lett.* **241**, 543 (2006).
4. L. Kennan, S. Lamb, C. Rundle, *J. S. Am. Earth Sci.* **8**, 163 (1995).
5. R. F. Kay, B. J. MacFadden, R. H. Madden, H. Sandeman, F. Anaya, *J. Vert. Paleontol.* **18**, 189 (1998).
6. B. J. MacFadden *et al.*, *J. Geol.* **93**, 223 (1985).
7. N. McQuarrie, *Geol. Soc. Am. Bull.* **114**, 950 (2002).
8. K. Elger, O. Oncken, J. Glodny, *Tectonics* **24**, 10.1029/2004TC001675 (2005).
9. Laboratory work reported in this response was supported by NSF grant EAR-0543952 to J.E.

14 August 2006; accepted 5 October 2006
10.1126/science.1133131

AGRICULTURE

Chewing on the Food Chain

David L. Katz

Six Arguments for a Greener Diet is an important book that deserves space on the shelf of anyone concerned about the well-being of their own person, other persons, other species, or the planet. Those meeting these criteria—I shudder to think of those who don't—should either square their shoulders and soldier through the book start to finish or at least keep it handy for reference whenever debate, deliberation, or dinner calls for well-packaged facts and a bracing reality check.

Six Arguments for a Greener Diet
How a More Plant-Based Diet Could Save Your Health and the Environment

by Michael F. Jacobson and the Staff of the Center for Science in the Public Interest

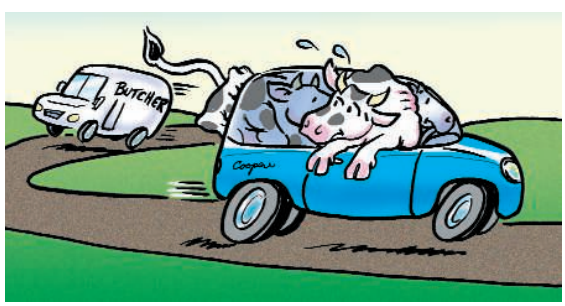
Center for Science in the Public Interest, Washington, DC, 2006. 248 pp. Paper, $14.95, C$21. ISBN 0-89329-049-1.

The book is neither beautiful nor fun. The authors' case for a more plant-based diet has the approximate literary flourish of a medical safety data sheet. But for literary indulgence, there is always Patrick O'Brian. *Six Arguments* is dissertation, not diversion. And as such, it offers the merits of thoroughness and careful research. Each of the six claims—that a diet with less meat consumption would reduce chronic disease; reduce foodborne illness; improve the quality of soil, water, and air; and reduce animal suffering—invokes numerous citations (no fewer than 50 to more than 220). The relentless factual barrage is clearly designed to leave readers devoid of resistance, if not breath.

But breathlessness here is more a matter of exhaustion than inspiration. Michael Jacobson and his colleagues (the authors are all at the Center for Science in the Public Interest, a nutrition-advocacy organization in Washington, DC) unambiguously want us all to eat less meat (or, preferably, none). In pursuit of this goal, the book plays like the percussion section of a marching band, a relentless drumbeat of data intended to get us all marching in step. The cadence is compelling, but ponderous and mostly quite predictable.

Still, surprise is not altogether absent. While discussing the regulatory authority intended to protect us from foodborne illness, Asher Wolf notes that the U.S. Department of Agriculture has jurisdiction over dehydrated chicken soup, whereas the Food and Drug Administration is responsible for dehydrated beef soup. However, the FDA oversees chicken broth and the USDA, beef broth. Laughs are even rarer than surprises in *Six Arguments*, but this tidbit of bureaucratic bungling was conducive to both.

Quantitative surprise comes along with greater frequency if less mirth. We learn that 15,000 to 20,000 gallons of water are consumed in the production of one pound of edible beef, and that the United States is depleting its underground aquifers by some 21 billion gallons per day. In 2000, methane produced by U.S. livestock contributed as much to global warming as the emissions of roughly 33 million automobiles. Under dairy industry recommendations that call for 20 to 25 square feet of space per 1000 pounds of animal, the footprint of the Mini Cooper automobile "would accommodate three adult cows with some room to spare."



CREDIT: JOE SUTLIFF

Similar facts have been served up before. More than three decades ago, Frances Moore Lappé addressed the links among dietary pattern, human health, and the environment in *Diet for a Small Planet* (*1*). More recently, John Robbins, heir to the Baskin-Robbins fortune he chose to renounce, gave us *The Food Revolution* (*2*). Lappé blended kitchen table wisdom with sociopolitical activism and was long on philosophy. Robbins used anecdotes to convey his pathos and evoke that of his reader. Jacobson has clearly opted for persuasion by the percussive force of factual onslaught. Perhaps it is upon a foundation of dispassionate fact that our collective verdict on this topic should be based. But while we get truth here, and almost nothing but, there is the occasional distortion and omission.

The whole truth would seem to require some consideration of why humans developed a substantially animal-food diet in the first place. Our intent was surely not to destroy our health or ravage the planet; those are unfortunate by-products of some other drive. Such a drive—native taste preferences, forged in the cauldron of natural selection, for instance—must be acknowledged to be overcome. But on the origins of our plight and its seductive, transcultural allure, Jacobson is silent.

As for distortion, the authors espouse what is fast becoming an obsolete view on the adverse health effects of dietary cholesterol, and they tar eggs and milk and meat with the same broad brush. Although there may be environmental reasons to eat fewer eggs, the health argument is increasingly cutting the other way (*3*). A willingness to acknowledge when personal, animal, and environmental health objectives part company would create a clearer impression of objectivity—and fortify the authors' arguments on those far more frequent occasions when they do, indeed, coincide.

Six Arguments does not come with a mission statement. If it did, I suspect that would be less concerned with enlightening the blissfully oblivious than with assaulting the ambivalent. The book's ideal audience seems to be those who are vaguely aware, and certainly care, but whose behavior does not quite correspond with their beliefs. We may nurture fantasies of Old MacDonald's farm, where animals could be singled out and doubtless all had names. Even those preoccupied with the grim environmental warnings in Al Gore's *An Inconvenient Truth* may fail to perceive any implications for global warming from the roast beef sandwich they eat while watching the film. Jacobson *et al.* will brook no such denial.

I am rooting for the authors and would like their arguments to carry the day. I am thoroughly persuaded that the health of people, and the planet, would improve were we all to eat a more plant-based diet. It seemed almost providential that as I was reading the book, an outbreak of *Escherichia coli* O157H7–tainted spinach was dominating news headlines. This pathogen, as the authors note, has emerged because reliance on feed grain has acidified the gastrointestinal tract of cattle; *E. coli* O157-H7 is relatively acid resistant. Regrettably, the news coverage focusing on the temporary dangers of spinach may have failed to convey, as

The reviewer is at the Prevention Research Center, Yale University School of Medicine, 130 Division Street, Derby, CT 06418, USA. E-mail: david.katz@yale.edu

the book does, that the contaminated produce is really just an innocent bystander of animal husbandry gone awry.

Chewing on the long-forged links in a culture-bound food chain is not for the faint of tooth. But nobody said it's easy being green—or greener, for that matter. The authors do claim that the fate of the world may depend on such an effort, and their arguments are hard to resist.

References

1. F. M. Lappé, *Diet for a Small Planet* (Ballantine, New York, 1971).
2. J. Robbins, *The Food Revolution: How Your Diet Can Help Save Your Life and the World* (Conari, Berkeley, CA, 2001).
3. D. L. Katz *et al.*, *Int. J. Cardiol.* **99**, 65 (2005).

10.1126/science.1134485

PLANT ECOLOGY

A Field Trip Through Diversity

Sara Alexander

Reading *Demons in Eden: The Paradox of Plant Diversity* is like listening to a great professor deliver a stellar lecture. The professor's passion for the subject is undeniable, the notes are well ordered, the explanations are clear and memorable, and the digressions are always illuminating, amusing, or both. The book also bears some resemblance to a travel novel, with chapters using rich metaphors and vivid descriptions to take the reader to locations that range from seemingly familiar to truly exotic. In addition, it serves as a call to action, marshaling findings from basic ecological research to paint a frightening picture of how much damage humanity has already done to diversity.

Demons in Eden
The Paradox of Plant Diversity
by Jonathan Silvertown
University of Chicago Press, Chicago, 2005. 187 pp. $25, £16. ISBN 0-226-75771-4.

The paradox mentioned in the book's title is this: If natural selection promotes those plants that are most successful in producing offspring, why has evolution resulted in the vast diversity we see today rather than only a few, hypersuccessful species that outcompete all others? Jonathan Silvertown, a plant ecologist at the Open University, takes the reader along on a journey around the world to find the answer. The engaging result is appropriate for a wide range of readers, even those who have only a very basic knowledge of biology.

Heather and gorse on an English heath.

Silvertown's term "Darwinian demon" has been used before, but with a different meaning. Richard Law introduced it to describe a hypothetical organism that "can maximize all aspects of fitness simultaneously"—that is, one that reproduces immediately after birth, has an infinite number of offspring, and lives forever (*1*). Taking a broader view, Silvertown sees as Darwinian demons those species that catapult "from obscurity to dominance in just a handful of generations, producing evolutionary change." He even comments that "all species must go through a demon phase in their evolutionary history." The demons envisioned by Law can never exist, and Silvertown's demons don't conquer every habitat for many of the same reasons, including limited resources, the need to balance the costs of growth and reproduction, competition with other species, and predation (*2*). The author doesn't simply tell the reader these conclusions; he shares every step of the process by which he reached them. Thus, after reading the book, one not only knows the answer but understands it.

Students and lay readers will also come away from the book with some idea of the fascinations and frustrations of basic research. Silvertown's enthusiasm for scientific sleuthing is infectious, but he does not shy away from the grittier aspects. He describes carefully planting seeds using forceps and painstakingly watching the seedlings for months, only to have his experiment end inconclusively. Silvertown quotes the 19th-century chemist Justus von Liebig, who in defending his incorrect claim that plants obtain their nitrogen from ammonia in the atmosphere called the experiments of his opponents "entirely devoid of value" and "most impudent humbug." He also notes the discouraging odds researchers face at high-prestige journals, where most manuscripts are rejected without being sent out for peer review.

Happily, Silvertown also highlights examples of brilliantly simple and rewarding studies and recounts success stories that would spark any aspiring biologist's imagination. Fir forests on the slopes of Mount Shimagare in Japan are naturally arranged not in patches of different ages, but in "waves" of progressing age, so a visitor can study 80 years of tree life history by walking a few dozen meters. A graduate student, Carly Stevens, found that acid grasslands in Britain are so sensitive to nitrogen air pollution that for every 2.5 kilograms of nitrogen deposited per hectare, one species disappears from the area's flora (*3*); her surprising discovery made news around the world. Silvertown also points out areas where future research is needed and provides examples of how new data can challenge old theories. Discussing the fate of Dan Janzen and Joseph Connell's enemies hypothesis (the idea that the concentration of natural enemies around adult trees subjects nearby offspring to fatal levels of attack), he shows how disproved theories, slightly changed, can turn out to be correct. He demonstrates that a scientist must always be ready to admit possible mistakes, examine assumptions, repeat results, and step back to examine even the most satisfying conclusion: "A good rule in science … is that if something is too good to be true, it probably isn't. So, we checked and double-checked our data, and then did the same kind of analysis for an even bigger dataset.… [W]e got the same result.… The next question was 'Why?'"

Although *Demons in Eden* covers a broad range of information, the author's conversational style and skill at concise, clear explanations keep him from alienating the layperson or boring the scientist. The book's length precludes a comprehensive discussion of plant diversity, so students will do best to pair Silvertown's engaging account with textbook study. Having finished the book just before starting a course in population ecology, I particularly appreciate the real-life illustrations it provides for many of the abstract topics covered in lecture.

References

1. R. Law, *Am. Nat.* **114**, 399 (1979).
2. M. B. Bonsall, V. A. A. Jansen, M. P. Hassell, *Science* **306**, 111 (2004).
3. C. J. Stevens, N. B. Dise, J. O Mountford, D. J. Gowing, *Science* **303**, 1876 (2004).

The reviewer is at the Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013–7012, USA. E-mail: alexandersar@si.edu

CREDIT: MIKE DODD

CORRECTED 22 DECEMBER 2006; SEE LAST PAGE

CLIMATE CHANGE

An Ambitious, Centrist Approach to Global Warming Legislation

A long-term declining cap on emissions, cost control, and strategic use of emission allowances to promote new technology would better curb global warming than current legislative proposals.

David D. Doniger, Antonia V. Herzog, Daniel A. Lashof*

There is growing concern that global warming of more than 2°C from preindustrial levels could have dangerous climatic consequences (*1*, *2*). It is estimated that, to avoid exceeding this 2° target, heat-trapping gas and aerosol concentrations need to be stabilized so that their net radiative effect is less than that of 450 parts per million (ppm) CO_2 (*3*). This could be achieved if the United States and other industrial nations cut current emissions by 60 to 80% by 2050, and if developing countries limit emissions growth and impose similar reductions later in the century.

In June 2005, the U.S. Senate passed a landmark resolution calling for a "comprehensive and effective national program of mandatory, market-based limits and incentives on emissions of greenhouse gases that slow, stop, and reverse the growth of such emissions" (*4*). A half-dozen legislative proposals have been introduced, and more are expected in 2007, all creating "cap-and-trade" systems like the successful program enacted in 1990 to curb acid rain.

Some bills would take only an initial step to slow or stop emissions growth, putting off decisions on future emission reductions for a decade or longer (*5*, *6*). Other bills would start with similar reductions, but would also set a longer-term emissions cap that declines at a predictable rate until 2050 (*7*, *8*).

The former proposals are based on the assumption that a more ambitious approach is not now politically feasible. In our view, we no longer have the time for such two-step strategies. Most climate scientists now warn that time is short for beginning serious emission reductions if we are to avoid dangerous climate impacts (*5*). A new approach is needed that is capable of garnering enough support to be enacted promptly while also requiring the deep reductions needed by mid-century.

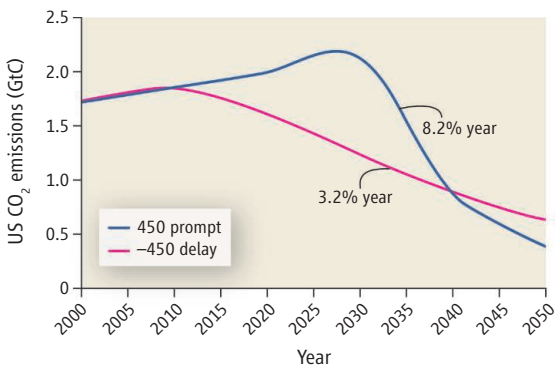
Sometimes decisions can be postponed without great cost; not so with global warming. Heat-trapping emissions are cumulative, and delaying the decision to reduce emissions will only worsen the problem and make the task of solving it much harder.

This is illustrated in the two hypothetical emission reduction scenarios for the United States presented below. Either scenario, in concert with comparable action by other nations, is aimed at avoiding atmospheric concentrations higher than 450 ppm CO_2 equivalent. But the two scenarios have vastly different economic implications.

If national emission reductions start soon, we can stay on the 450-ppm path with an annual emission reduction rate that gradually ramps up to 3.2% per year. But if we delay a serious start by, for example, 20 years and allow continued emission growth at nearly the business-as-usual rate, the annual emission reduction rate required to stay on this path jumps to 8.2% per year (see the figure).

Some analysts argue that delay is cheaper, because we will develop breakthrough technologies in the interim. But that outcome is implausible for three reasons. First, delay dramatically increases the emission reduction rate required later. Cutting emissions by more than 8% per year would require deploying advanced low-emission technologies several times faster than conventional technologies have been deployed over recent decades (*9*). Second, without meaningful near-term market signals, there will be little incentive for the private sector to direct significant R&D resources toward developing the breakthrough technologies. Hope will rest entirely on the federal R&D program, which now is far too small to yield the required results. Third, without market signals, a new generation of conventional power plants, vehicles, and other infrastructure will be built during the next two decades. Our children and grandchildren will have to bear the costs of prematurely retiring an even bigger stock of highly emitting capital than exists today. Even with a substantial discount rate, it is virtually impossible that delaying emission reductions will be cheaper than starting now. These factors argue that making an early start is the most effective way to minimize the overall economic impact of the necessary emission reductions.

We need a formula that will reduce global warming emissions sufficiently and still meet the legitimate economic concerns of industries and other constituencies. The key elements of our formula include the following.



A slow start leads to a crash finish. Prompt-start and delay pathways consistent with stabilizing heat-trapping gases at 450 ppm CO_2 equivalent. Global emissions 2000–2100 are 480 gigatons carbon (GtC) from Meinshausen's S450Ce scenario (*14*). The U.S. share of global emissions is assumed to decline from 25 to 5% linearly between 2000 and 2100. This results in an emissions budget of 84 GtC in the 21st century. In the prompt-start case, emissions decline by 1.5% per year from 2010 to 2020, 2.5% per year from 2020 to 2030, and 3.2% per year thereafter. The delay case assumes that emissions grow by 0.7% per year from 2010 to 2030, a reduction of 0.5% annually compared with the Energy Information Administration forecast (*15*); emissions must decline by 8.2% per year thereafter to limit cumulative 21st-century emissions to 84 GtC. Cumulative emissions 2000–2050 are 68 GtC in the prompt-start scenario and 79 GtC in the slow-start scenario.

A prompt start and long-term declining cap. The first element of an effective climate protection bill is to put U.S. emissions on a pathway consistent with a prompt-start and declining cap as shown in the figure. This pathway requires earlier action than many in the business community now expect, but it also provides them important economic benefits. It will create the certainty needed for efficient planning of long-lived capital investments. It will also be less costly and more predictable than a pathway dependent on crash reductions later on. The possibility of revisiting and fine-tuning the long-term target intro-

The authors are with the Natural Resources Defense Council, Washington, DC 20005, USA.

*Author for correspondence. E-mail: dlashof@nrdc.org

duces much less uncertainty than if the long-term target is left entirely undefined for another decade or more.

A new approach to controlling unexpected costs. Long-term costs are not industry's only concern. The greater fear for many is that costs will fluctuate unexpectedly in the short-run, much as natural gas prices have spiked in recent years. Setting a long-term emissions cap opens the door to an innovative way to avoid short-term cost volatility: Firms could be allowed to borrow emissions allowances from future years, using them early in times of unexpected cost pressure, and paying them back when short-term spikes recede.

In cap-and-trade programs, each firm must turn in an emissions allowance for each ton of pollution that it emits. Total allowances are limited to the number of tons allowed by the cap. Current legislative proposals already allow firms to make reductions in advance when prices are lower than expected and to bank allowances for future use. Borrowing would open the opposite possibility.

In the absence of borrowing, firms can comply only with current or banked allowances. Allowance prices thus reflect the current marginal cost of compliance, and that price can spike in response to short-term conditions (e.g., a delay in bringing on a new technology, or a surge in economic activity). Borrowing will let firms use emissions allowances from future years, stabilizing prices against unexpected short-term fluctuations. The long-term cap will be maintained, because borrowed allowances will be repaid, with interest, by releasing fewer emissions later when short-term pressures are relieved.

The combination of a long-term emissions pathway and borrowing has advantages over other cost-control proposals that break the cap and permanently allow excess emissions. Under the "safety-valve" proposal, for example, the government sells more allowances if the price per ton exceeds a designated level. Under one proposal, the safety-valve price would start at $7 per ton of CO_2, a level virtually certain to be exceeded and that would result in rising emissions at least through 2020 (*5*, *6*, *10*). Although the safety valve may close in future years, the excess emissions that occur while the safety valve is open will never be recouped. Likewise, proposals allowing unlimited "offsets" (emission reductions not covered by the cap) have the potential to break the cap if credits are awarded for actions taking place anyway—a problem endemic to past offset programs.

Promoting technology. Emissions allowances will be worth hundreds of billions of dollars over the life of the program. We propose that allocation of emissions allowances be used strategically, in combination with targeted performance standards, to encourage faster deployment of low-carbon technologies, and for other purposes.

It is a common misconception that regulated companies will be grievously hurt unless they receive all the emission allowances they need free of charge. In reality, firms can be expected to pass most compliance costs on to consumers, and only a fraction of those costs will fall on shareholders. If regulated industries got all their allowances free, they would receive an asset worth as much as seven times the real cost of compliance (*11*), resulting in substantial windfall profits, as has been seen in Europe (*12*). As a result, Congress has begun a serious discussion of how allowances should be allocated (*13*).

We recommend allocating half of the allowances to helping businesses and consumers (particularly energy-intensive industries and lower-income families) reduce their energy bills through adopting currently available energy-saving technology and competitive renewable energy sources, such as wind power.

Another quarter of the allowances should be allocated to companies that accelerate deployment of strategic new technologies needed for long-term emission reductions in key sectors. Incentives should focus on rebuilding the electric power industry by means of advanced fossil fuel technologies equipped with geologic CO_2 disposal and advanced renewable energy technologies, retooling the auto industry to make more hybrids and other low-emitting vehicles, and jump-starting farm production of biofuels from cellulosic feedstocks.

This approach would provide a stable long-term source of funding to advance key technologies without increasing the federal deficit. It would also build political support in critical economic constituencies.

The remaining emission allowances can be allocated to meet other key needs. For example, revenue from a share of the allowances could be used to help communities heavily affected by climate impacts or heavily impacted by mitigation measures. Some allowances could also be used to encourage carbon sequestration in our soils and forests.

Rejoining the World

For too long, the United States has excused its own inaction by saying that it cannot solve this problem alone. But other countries cannot be expected to play their full part if the world's largest emitter continues to do nothing. Global progress requires that we begin to act at home and rejoin international negotiations with a new attitude. As the Senate resolution states, by beginning to curb our own emissions we will encourage others to act (*4*). At the same time, it is reasonable to take stock periodically to ensure that others are taking reciprocal action. Several current legislative proposals usefully provide for a regular review every 5 or 10 years based on input from the Administration and the National Academy of Sciences on the current science, economics, and state of international cooperation (*6*–*8*). Congress would then decide whether or not to fine-tune the declining cap.

We think this package, or some variation, has the potential to begin bridging the gap between environmental and business advocates and to build centrist, bipartisan support for effective climate legislation.

References and Notes

1. J. Hansen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14288 (2006).
2. H. J. Schellnhuber, W. Cramer, N. Nakicenovic, T. Wigley, G. Yohe, Eds. *Avoiding Dangerous Climate Change* (Cambridge Univ. Press, New York, 2006).
3. M. Meinshausen, in (*2*), pp. 265–279.
4. Senate Amendment 866 to the Energy Policy Act of 2005 (H.R. 6), A sense-of-the-Senate climate change resolution. *Congress. Rec.* **151** (22 June 2005), S7033–S7037, S7089.
5. National Commission on Energy Policy, *Ending the Energy Stalemate: A Bipartisan Strategy to Meet America's Energy Challenges* (National Commission on Energy Policy, Washington, DC, 2004).
6. Senate Amendment 868 to the Energy Policy Act of 2005 (H.R. 6), "Climate and Economy Insurance Act of 2005," *Congress. Rec.* **151** (22 June 2005), S7090–S7098.
7. Senate Bill S. 3698, "Global Warming Pollution Reduction Act," 18 July 2006.
8. House of Representatives Bill H.R. 5642, "Safe Climate Act of 2006," 20 June 2006.
9. D. Hawkins, in *Greenhouse Gas Control Technologies: Proceedings of the 7th International Conference on Greenhouse Gas Control Technologies* (*GHGT7*), M. Wilson *et al.*, Eds., 5 to 9 September 2004, Vancouver, Canada (Elsevier Science, Kidlington, Oxford, UK, 2005), pp. 1525–1530.
10. Energy Information Administration (EIA), U.S. Department of Energy (DOE), *Impacts of Modeled Recommendations of the National Commission on Energy Policy* (SR/OIAF/2005-02, DOE, Washington, DC, 2005).
11. D. Burtraw, K. Palmer, D. Kahn, "Allocation of CO_2 emission allowances in the regional greenhouse gas cap-and-trade program" (Discussion paper, Resources for the Future, Washington, DC, June 2005).
12. J. Sijm, K. Neuhoff, Y. Chen, *Climate Pol.* **6**, 49 (2006).
13. Senators Pete V. Domenici (R–NM) and Jeff Bingaman (D–NM), "Design elements of a mandatory market-based greenhouse gas regulatory system" (Climate change white paper, February 2006), available at http://energy.senate.gov/public/_files/ClimateChangeWhitePaper.doc.
14. Simple Model for Climate Policy assessment (SiMCaP), available at www.simcap.org/.
15. Reference case from (DOE), *Annual Energy Outlook 2006 with Projections to 2030* (Report no. DOE/EIA-0383, DoE, Washington, DC, 2006).

10.1126/science.1131558

CELL BIOLOGY

Nuclear Pore Complex Models Gel

Two views are given on the elastic structure of pores in the cell's nuclear membrane, which allows the exchange of materials between the nucleus and the cytoplasm.

Brian Burke

In eukaryotic cells, the boundary between the nucleus and cytoplasm is defined by a membranous organelle, the nuclear envelope. Trafficking of macromolecules back and forth across this envelope occurs through nuclear pore complexes (*1*). A vertebrate somatic cell typically contains between 1000 and 10,000 such pore complexes. Small molecules can pass through these pores unimpeded; larger molecules (30- to 40-kD proteins) must associate with soluble nuclear transport receptors and be escorted through the central channel of the pore. We know much about the structure of the nuclear pore complex and the role of soluble components in nucleocytoplasmic transport, but the mechanics of translocation through the nuclear pore complex have been debated. On page 815 of this issue, Frey *et al*. (*2*) provide evidence for the existence of a flexible sieve that spans the pore, creating a selective permeability barrier.

Nuclear pore complexes are composed of multiple copies of about 30 different protein subunits (nucleoporins or nups). About one-third of these nups contain FG domains featuring arrays of the hydrophobic peptide repeats FG, GLFG, or FXFG (F, Phe; G, Gly; L, Leu; X, any amino acid). FG domains are thought to be natively unfolded, adopting extended, flexible conformations. They are also considered to have a central role in nuclear pore complex function, because FG repeats bear binding sites for nuclear transport receptors.

But how exactly do FG nups mediate nucleocytoplasmic transport? Two mechanistic models have dominated this discussion. The first, proposed by Rout *et al*. (*3*), invokes the concept of virtual gating. In this scheme, FG nups increase the residence time of transport complexes in the central aperture of the pore by binding to nuclear transport receptors. In this way, FG nups facilitate diffusion of transport complexes into the central channel. Conversely, because FG domains are flexible and largely unstructured, they limit available space in the nuclear pore complex near-field, thus restricting access of nontransport substrates to the nuclear pore complex. Recently, Aebi and colleagues (*4*) used atomic force microscopy to study the FG domain of vertebrate Nup153 immobilized on gold nanodots. They concluded that FG domains cluster and form a "polymer brush" that could indeed display the type of exclusionary function that is key to virtual gating. A more recent proposal based on constrained diffusion within the nuclear pore complex central channel also highlights aspects of virtual gating (*5*).

Ribbeck and Görlich (*6*) proposed an alternative model in which FG domains within the central channel of the pore complex interact through FG repeats to form a protein meshwork, essentially forming a separate hydrophobic phase. Transport complexes can partition into this phase because of their capacity to bind to the FG repeats, thereby locally perturbing FG domain interactions. Proteins incapable of binding FG repeats are excluded from this hydrophobic phase. Small molecules and proteins below the size exclusion limit pass through the interstices of such a meshwork, independent of transport receptors. An earlier proposal by Macara and colleagues, called the "oily spaghetti" model (*7*), presents some features of the selective phase concept and similarly underscores the hydrophobic nature of the unstructured FG domains.

An important difference between the virtual gating and selective phase models concerns the interaction between FG repeats.

The author is in the Department of Anatomy and Cell Biology, University of Florida College of Medicine, 1600 SW Archer Road, Gainesville, FL 32606, USA. E-mail: bburke@ufl.edu

MATERIALS SCIENCE

Polymers in the Pore

Michael Elbaum

The nuclear pore complex may be nature's ultimate analytical chemist. Seated at the gateway between the nucleus and cytoplasm in eukaryotic cells, it distinguishes a mixed solution of macromolecules by their chemical identity, all the while remaining open to diffusive passage of water, ions, metabolites, and other small solutes. From a physical point of view, it is a fascinating machine. Frey *et al.,* on page 815 in this issue, explore the unlikely talent of this specially tuned barrier (*1*).

Up to a size cutoff of a few nanometers, the nuclear pore acts as a simple sieve. Beyond ~40 kD, most proteins and protein complexes are unable to cross it on their own. Nuclear transport receptors may usher such larger cargoes specifically across the pore. Ironically, a midsize protein must recruit a large receptor to pass through a narrow channel. Clearly there is something special in the recognition of transport receptors by whatever makes up the sieve within the pore. Attention has focused on repeat motifs of phenylalanine-glycine (FG) that are common among constituent proteins of the nuclear pore itself. The FG repeats do indeed interact with nuclear transport receptors. Moreover, these domains tend to be natively unfolded polypeptides, so they are presumed to swell into the central channel of the pore. From here, the key questions are essentially of polymer physics and chemistry.

What is the nature of this FG-repeat network? It was proposed to form a hydrogel, cross-linked by hydrophobic interactions between the phenylalanines (*2*). Frey *et al*. show that FG repeats can indeed form a free-standing gel, and they measure elasticity comparable to 0.4% agarose. They show further that mobility of fluorescently labeled FG peptides is low in the gel, consistent with cross-linking among them. Mutating the phenylalanines to serines results in both loss of gel stiffness and a higher mobility of the polymers.

A natural scale for the nuclear pore sieve is then simply the mesh size of the gel. Mobility of

The author is in the Department of Materials and Interfaces, Weizmann Institute of Science, 76100 Rehovot, Israel. He is presently on sabbatical at Max Planck Institute of Colloids and Interfaces, Potsdam, Germany. E-mail: michael.elbaum@weizmann.ac.il

The former model predicts mobile FG domains, implying minimal interactions. This prediction is supported by the atomic force microscopy observations of the Nup153 FG domain. The latter model requires more stable interactions between FG domains to establish a meshwork. Frey *et al*. have found that the FG domain of yeast Nsp1p will form a mechanically stable elastic hydrogel in vitro, featuring hydrophobic interactions between FG repeats. These Nsp1p hydrogels can even incorporate the FG domains of other FG nups. Hydrogel formation is absolutely dependent on the FG repeats because it is abolished by substitution of S (Ser) for every F within the Nsp1p FG domain (Nsp1pF→S).

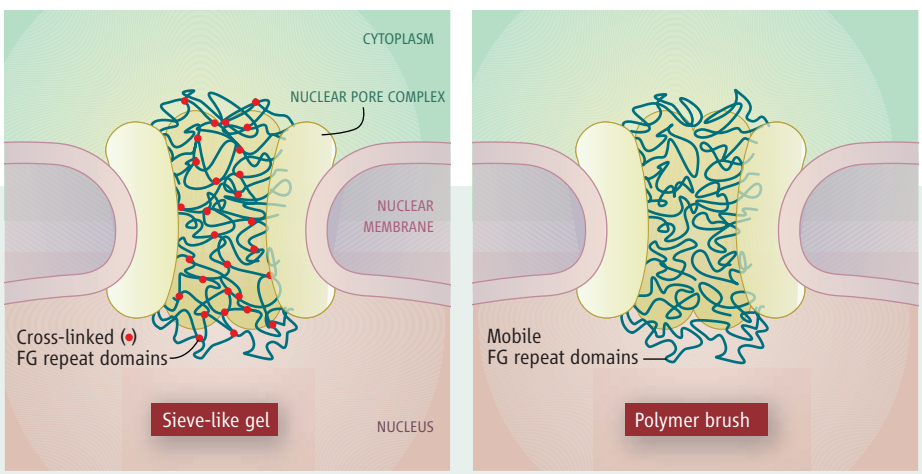
Wente and colleagues had previously demonstrated redundancy in the FG domains of yeast nups (*8*). In particular, FG domains of nups that are distributed asymmetrically at the nucleoplasmic or cytoplasmic face of the nuclear pore complex appear dispensable. Indeed, deletion of the FG domain of Nsp1p, which localizes to the cytoplasmic side of the yeast pore complex, has little or no effect on viability. However, Frey *et al*. now show that substitution of Nsp1p by Nsp1pF→S is lethal in yeast and that this lethality cannot be attributed to inability to bind nuclear transport receptors. They suggest that Nsp1pF→S perturbs the hydrophobic characteristics of the FG phase within the nuclear pore complex, leading to loss of pore functionality.

This new study provides some compelling evidence for the selective phase model. Nonetheless, questions still remain. For instance, the FG domains of certain vertebrate nups are extensively modified with O-linked *N*-acetylglucosamine (*9*). How this might affect interactions between FG domains remains to be seen. Furthermore, not all FG domains may be equivalent. A distinct possibility is that while certain nups may contribute to a selective FG phase within the core of the pore complex, others at the periphery might behave more like a virtual gate. As biophysical and genetic approaches are brought to bear on the problem, the answer may be close at hand.

References

1. E. J. Tran, S. R. Wente, *Cell* **125**, 1041 (2006).
2. S. Frey, R. P. Richter, D. Görlich, *Science* **314**, 815 (2006).
3. M. P. Rout, J. D. Aitchison, M. O. Magnasco, B. T. Chait, *Trends Cell Biol.* **13**, 622 (2003).
4. R. Y. Lim *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9512 (2006).
5. R. Peters, *Traffic* **6**, 421 (2005).
6. K. Ribbeck, D. Görlich, *EMBO J.* **21**, 2664 (2002).
7. I. G. Macara, *Microbiol. Mol. Biol. Rev.* **65**, 570 (2001).
8. L. A. Strawn, T. Shen, N. Shulga, D. S. Goldfarb, S. R. Wente, *Nat. Cell Biol.* **6**, 197 (2004).
9. G. D. Holt *et al*., *J. Cell Biol.* **104**, 1157 (1987).

10.1126/science.1135739



Pore gel. Alternate views on the polymeric state of natively disordered FG-repeat domains swelling into the transport channel of the nuclear pore.

nuclear transport receptors is explained in this scheme by their ability to replace the labile bonds between the polymers with links to themselves. In other words, they dissolve into the gel. Partitioning of transport receptors between the gel and surrounding aqueous phases allows for molecular exchange and transport. Just as hydrophobic moieties cross lipid bilayers much faster than hydrophilic ones, specific hydrophobic interactions between the FG proteins and the transport receptors allow the receptors to cross the FG sieve as a "selective phase" barrier (*2*). Frey *et al.* show that mutating the phenylalanines to tyrosines suppresses the interaction with transport receptors, but the gel state is retained, showing that the two features are independent.

An alternate view holds that FG repeats could form a network of unlinked polymers whose thermally activated undulations create a zone of "entropic exclusion" (*3*). The principle is similar to stabilization of colloids by capping their surfaces with long-chain molecules. The entropic penalty in collapsing these chains prevents aggregation of neighboring particles. By transiently attaching to the FG polypeptides, perhaps at multiple points, transport receptors could circumvent this exclusion. Conceptually, this model is inspired by weak repulsive forces between neurofilaments, a cytoskeletal structure that gives mechanical strength to axons and dendrites in neurons (*4*). Indeed, Lim *et al.* (*5*) recently found that end-anchored FG repeats show entropy-dominated elastic properties of a "polymer brush" (*6*). The force measured in compressing the brush grows exponentially as the gap is closed. At least in vitro, the FG-repeat networks tested can take both proposed forms.

Probing the mobility of nuclear transport receptors in well-defined FG gels or brushes will require further analyses, catching up in a way with single-molecule studies made recently in native nuclear pores (*7*, *8*). Whichever mechanical model of its sieve turns out to be more relevant in the cellular context, understanding the polymer physics of the nuclear pore may inspire novel biomimetic materials or nanotechnological devices to corral specific macromolecules from a mixed solution. Enantiomeric separation by antibody (*9*) or polypeptide-lined (*10*) membrane pores could make an interesting start in this direction. A long road lies ahead, though, until materials science can match the exquisite single-residue sensitivity of the nuclear pore and its transport receptors.

References

1. S. Frey, R. P. Richter, D. Görlich, *Science* **314**, 815 (2006).
2. K. Ribbeck, D. Görlich, *EMBO J.* **20**, 1320 (2001).
3. M. P. Rout *et al*., *J. Cell Biol.* **148**, 635 (2000).
4. H. G. Brown, J. H. Hoh, *Biochemistry* **36**, 15035 (1997).
5. R. Y. H. Lim *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9512 (2006).
6. H. J. Taunton, C. Toprakcioglu, L. Fetters, J. Klein, *Macromolecules* **23**, 571 (1990).
7. U. Kubitscheck *et al*., *J. Cell Biol.* **168**, 233 (2005).
8. W. Yang, S. M. Musser, *J. Cell Biol.* **174**, 951 (2006).
9. S. B. Lee *et al*., *Science* **296**, 2198 (2002).
10. N. H. Lee, C. W. Frank, *Polymer* **43**, 6255 (2002).

10.1126/science.1135924

MATERIALS SCIENCE

Colloidal Encounters: A Matter of Attraction

Colloidal particles are often viewed as large-scale models for molecules, but they can show counterintuitive phase behavior.

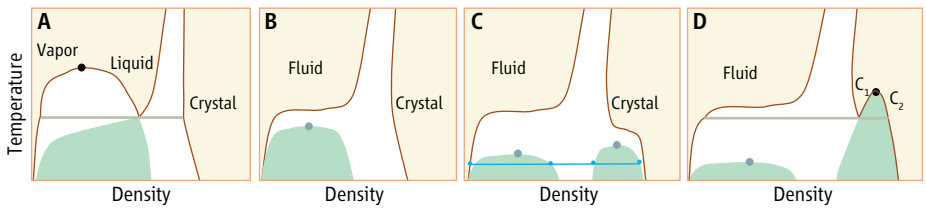
Daan Frenkel

In 1908, Heike Kamerlingh Onnes was the first to liquefy helium. Years of hard work preceded this breakthrough, but Kamerlingh Onnes was confident that the experiment was possible. In this he was guided by the principle of corresponding states formulated by Johannes van der Waals, which basically says that the relation among temperature, pressure, and density of all atomic and simple molecular fluids is similar—if one can be liquefied, then so can they all. On page 795 of this issue, Savage *et al*. (*1*) report experiments that reveal a complete breakdown of the corresponding-states principle.

Molecules repel each other at short distances but attract each other at longer distances. One can visualize this by considering a molecule as a castle (the harshly repulsive core) surrounded by a moat (the attractive region). The principle of corresponding states can be rationalized if we assume that molecules of different sizes are all scale models of the same design. It should even apply to suspensions of colloidal particles that are a thousand times the size of atoms. If these colloidal particles were scale models of atoms—that is, if they had the same ratio between the diameter of the hard core and the width of the attractive well—then colloidal suspensions would exhibit exactly the same phases (solid, liquid, and vapor) as their atomic counterparts (see the first figure, panel A).

However, the forces between colloids can be very different from those between atoms; the fact that scientists can tailor these forces is one of the main reasons why the study of colloidal systems is such an exciting field. How, then, does the colloidal phase diagram change from the corresponding-states case when we make the attractive well narrower relative to the hard-core diameter?

First, the critical point (where liquid and vapor become indistinguishable) moves toward the triple point (below which the liquid phase is no longer stable). When the effective width of the attractive well becomes narrower than 15% of the hard-core diameter, the liquid-vapor transition disappears (see the first figure, panel B). Such colloidal systems have only two stable phases: fluid and crystal. Kamerlingh Onnes was lucky: The effective width of the attractive well of helium is larger than 15% of its diameter.

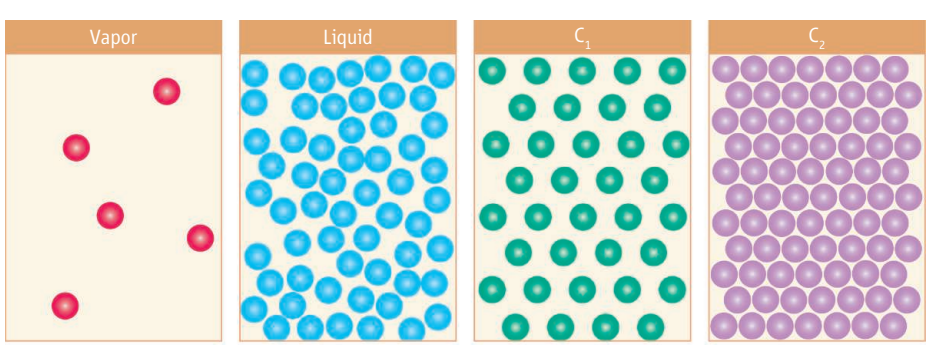


Tuning attraction. The temperature-density phase diagram of a system of spherical colloids changes when the range of the attractive forces between the particles is decreased. The solid curves are the phase boundaries of the stable phases, the dashed curves those of the metastable phases. Between the solid curves, two phases coexist. (**A**) Long-range attraction (the standard case for corresponding states). Solid circle, liquid-vapor critical point; gray bar, triple-point temperature. (**B**) Shorter-range attraction (less than 15% of the hard-core diameter). The liquid-vapor critical point (gray circle) has moved to the metastable region below the freezing curve. (**C**) Very-short-range attraction (2% of the hard-core diameter), but with a 3.5% spread in colloid size. In the metastable region, a solid-solid transition appears at high densities and a liquid-vapor transition at lower densities. The experiments of Savage *et al*. illustrate this case. The dissolution of a dense crystallite follows the horizontal blue line from right to left: First, the dense crystal transforms into an expanded crystal; the latter changes into a liquid phase that then evaporates to the stable vapor phase. (**D**) Very-short-range attraction with narrow particle-size distribution. A transition between an expanded (C_1) and a condensed (C_2) crystal phase ends in a critical point (solid black circle).

There are many experimental examples of colloidal systems with short-range attractions that have only two stable phases: fluid and crystal (*2*). For colloids with an even narrower attractive well (<5% of the hard-core diameter), simulations predict (*3*) two solid phases that differ only in density (see the second figure); above a critical temperature that may be either in the metastable (first figure, panel C) or the stable regime (first figure, panel D), these solid phases become indistinguishable. But until recently, it was not possible to tailor the colloidal systems needed to explore this behavior.

Savage *et al*. now report videomicroscopy experiments that overcome some of these problems. They have managed to prepare a system of spherical colloids with very-short-range attractive interactions: The width of the attractive well is less than 2% of the colloidal diameter.

In these experiments, the authors can vary the strength of the attraction between the colloids. Initially, the colloids form two-dimensional crystals on the coverslip of the sample container. These crystals coexist with a dilute colloidal "gas." Weakening the attraction between the colloids makes the crystals thermodynamically unstable. What then happens is very different from the melting of normal,



Colloidal phases. Sketch of the structure of the vapor, liquid, expanded crystal (C_1), and condensed crystal (C_2) phases. Vapor and liquid are fluid phases that differ only in density.

The author is at the FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098SJ Amsterdam, Netherlands. E-mail: frenkel@amolf.nl

three-dimensional crystals that tend to melt from the surface inward (*4*): The colloidal crystals slowly evaporate down to a critical size and then "explode" to form a dense amorphous phase. This phase is unstable and subsequently evaporates.

Why would a small colloidal crystal suddenly convert into an amorphous phase that is thermodynamically unstable? The answer may be contained in a paper published almost 40 years ago (*5*). There, Cahn explained how metastable phases can act as crucial intermediates during the transformation from an unstable to a stable phase. In the present experiments, the intermediate phase is the dense liquid that, for colloids with short-range attraction, is not thermodynamically stable. Evans and colleagues (*6*) have reported experimental evidence for the Cahn scenario in colloidal systems with a somewhat longer-range attraction.

Savage *et al*. do not show direct evidence for the solid-solid transition expected for systems with short-range attraction. But they show something else: As the sublimating crystals get smaller, their density decreases. This seems strange, because small liquid droplets tend to be more compressed than larger droplets as a result of the surface tension. However, the observed expansion of small crystallites agrees well with simulations (*7*), which predict such an effect in the vicinity of a metastable solid-solid critical point (see the first figure, panel C). Hence, although the solid-solid critical point in colloids has not yet been observed directly, it seems to be within reach. To observe it, the size distribution of the colloids would have to be narrower (see the first figure, panel D).

Many proteins have short-range attractions similar to those of the colloids, but their clustering and dissolution cannot be studied directly by videomicroscopy. Hopefully, microscopy studies of colloidal model systems such as those studied by Savage *et al.* will provide insights into the kinetics of phase transformations in these important biological processes.

References

1. J. R. Savage *et al.*, *Science* **314**, 795 (2006).
2. H. N. W. Lekkerkerker *et al.*, *Europhys. Lett.* **20**, 559 (1992).
3. P. Bolhuis, D. Frenkel, *Phys. Rev. Lett.* **72**, 2211 (1994).
4. J. W. M. Frenken, J. F. van der Veen, *Phys. Rev. Lett*. **54**, 134 (1985).
5. J. W. Cahn, *J. Am. Ceram. Soc*. **52**, 118 (1969).
6. F. Renth *et al.*, *Phys. Rev. E* **64**, 031402 (2001).
7. A. Cacciuto, S. Auer, D. Frenkel, *Phys. Rev. Lett*. **93**, 166105 (2004).

10.1126/science.1135544

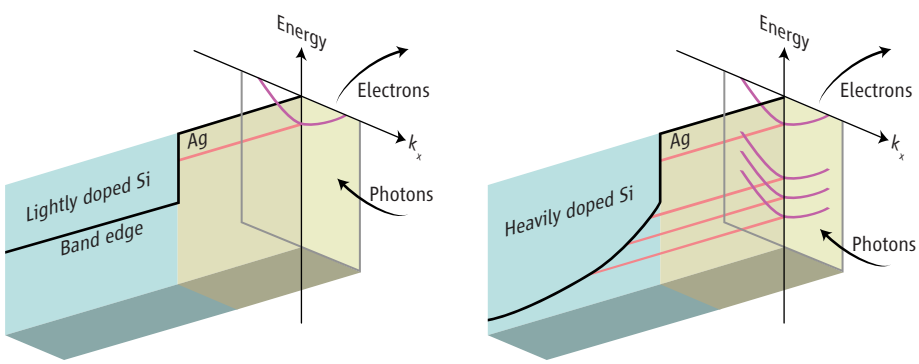
PHYSICS

Beyond the Particle in the Box

After a controversy-filled history, a simple explanation is provided for a new type of standing-electron-wave state observed in metal-semiconductor junctions.

Lars Walldén

Metal films and metal-semiconductor junctions are key components in modern electronic devices. On page 804 of this issue, Speer *et al.* (*1*) examine the energies of electrons in perfectly smooth and ultrathin silver films on silicon substrates. As expected, the authors observe energies that reflect the wave nature of electrons confined to a thin film. However, they also detect energies characteristic of a new type of electron wave. The latter extends through the film and into the substrate to a depth determined by the doping level of the substrate (see the figure). This observation has important implications regarding the basic physics of metal-semiconductor systems, especially those in which the film and the substrate have crystal lattices that do not match.



A matter of doping. Photons eject electrons from silver films deposited on silicon substrates, thereby yielding information about the electronic structure of the system. (**Left**) A single electronic level (or quantum-well state) is confined by the band gap of a lightly doped n-type silicon substrate. (**Right**) For a heavily doped n-type silicon substrate, the band edge of silicon varies near the interface, giving rise to additional quantum-well states that coherently span the silver film and a portion of the silicon substrate. In both cases, each quantum-well state (purple lines) exhibits a parabolic dispersion as a function of the in-plane momentum k_x.

The principles of quantum mechanics are usually introduced to students through the example of the particle in a box. In this example, a particle (such as an electron) is placed in a confining structure. The rules of quantum mechanics then specify the allowed states of the particle. An excellent example of particle-in-a-box (or quantum-well) behavior involves valence electrons in thin metal films. If the film is a simple metal such as aluminum, an alkali metal, or a noble metal, the electrons can essentially be regarded as free particles (but they are prevented from leaving the solid by a potential barrier at the surface). This confinement perpendicular to the surface leads to a set of standing waves, allowing only specific wavelengths to exist. In the parallel direction, there is no similar restriction on wavelength. The electronic structure is thus characterized by a series of electron energy bands, one for each wavelength allowed by the confinement.

The discrete character of the bands leads to film thickness–dependent properties (*2*). For example, the cohesive energy and the work function can show an oscillatory thickness dependence of substantial amplitude. These oscillations occur because, as the film thickness increases, the quantum-well states shift to lower energies to accommodate more electrons. At regular thickness intervals, new states become populated, such that the balance between electrons with high and low energy varies periodically with thickness. In experiments, the thickness is not a continuous variable but varies in steps given by the thickness of an atomic layer. If the period is incommen-

The author is in the Department of Applied Physics, Chalmers, 41296 Göteborg, Sweden. E-mail: wallden@fy.chalmers.se

surate with the layer thickness, one is left with a beat period. This beat period appears in transport properties such as magnetoresistance (*3*).

The discrete electronic structure of a thin metal film was first observed experimentally via the tunnel current across a metal-insulator thin-film sandwich (*4*). Because the film served as an electrode, it had to be conducting, which meant that the thickness had to be at least ~10 nm. Later work demonstrated that photoelectron (*5*, *6*) and scanning tunneling spectroscopies (*7*) are powerful tools to probe the electron states in thin adsorbed films down to monolayer thicknesses.

When the adsorbed film is only few atomic layers thick, the boundary conditions are particularly important. If a film is taken from vacuum and placed on a substrate, then it matters what the substrate is. Within a substrate energy gap (an energy range in which no valence electrons are found and which is often referred to as "forbidden"), one may find discrete quantum well–like states in the film. These states extend into the substrate, but only with an oscillating tail, with the period given by the substrate lattice. The decay depth of the state is given by the energy of the state with respect to the edges of the energy gap, the tails being long near the edges. The electronic state is thus a hybrid with a distinctly different character in the film and in the substrate.

Speer *et al.* have now used photoemission spectroscopy to study quantum-well states in silver films deposited on silicon. The silver films are 8 to 12 atomic layers thick. At these thicknesses, one would not expect the choice of substrate to be important for the ladder of energy bands. For example, nearly the same quantum-well state energies are observed when silver films are adsorbed on gold (*8*) as on the silicon substrate used by Speer *et al*. But as Speer *et al*. show, a dramatically different set of states can be obtained by increasing the doping of the silicon substrate.

When the Ag film and n-type Si are brought in contact, equilibrium (coincident Fermi levels) requires a transfer of electrons from Si to the metal. Near the interface, the semiconductor is depleted of electrons and the electron states are shifted in energy with respect to the states in the bulk. The depth dependence of the energy is referred to as band bending (*9*). With high n-doping (see the figure, right), the band bending saturates at a depth that is shallow enough for a novel set of discrete states to form. The electrons that form these states encounter the confining substrate band gap within the substrate and not at the interface. This means that the quantum well becomes wide, ranging from the vacuum interface to the depth where the energy coincides with the lower edge of the gap. The states therefore become less separated in energy than the standard type of quantum-well states.

The results reported by Speer *et al.* provide a detailed image of the electronic structure at a metal-semiconductor junction. The authors can account for their observations with simple models that can easily be extended to overlayers with different thickness and to substrates with different impurity content, or where the band bending may be changed by illuminating the interface (*10*).

Given the long and troublesome history of accounting for the properties of metal-semiconductor junctions (*11*), it is encouraging that there are cases in which simple modeling does not appear to be hampered by the occurrence of defects, intermixing, or compound formation at the interface. Furthermore, the effects reported by Speer *et al.* could be used to systematically modify the quantized electronic structure of thin film systems, thereby providing a powerful means for tuning properties of interest.

References

1. N. J. Speer, S.-J. Tang, T. Miller, T.-C. Chiang, *Science* **314**, 804 (2006).
2. F. K. Schulte, *Surf. Sci*. **55**, 427 (1976).
3. J. E. Ortega, F. J. Himpsel, G. J. Mankey, R. F. Willis, *Phys. Rev. B* **47**, 1540 (1993).
4. R. C. Jaklevic, J. Lambe, *Phys. Rev. B* **12**, 4146 (1975).
5. T.-C. Chiang, *Surf. Sci. Rep.* **39**, 181 (2000).
6. S. Å. Lindgren, L. Walldén, in *Handbook of Surface Science, Vol. 2, Electronic Structure,* S. Holloway, N. V. Richardson, K. Horn, M. Scheffler, Eds. (Elsevier, Amsterdam, 2000), pp. 89–95.
7. C. Corriol *et al.*, *Phys. Rev. Lett.* **95**, 176802 (2005).
8. T. Miller, A. Samsavar, G. E. Franklin, T.-C. Chiang, *Phys. Rev. Lett*. **61**, 1404 (1988).
9. C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, ed. 8, 2005), p. 506.
10. K. Horn, M. Alonso, R. Cimino, *Appl. Surf. Sci.* **56–58**, 271 (1992).
11. R. T. Tung, *Mater. Sci. Eng*. **R35**, 1 (2001).

10.1126/science.1135737

ECOLOGY

A Renaissance in the Study of Abundance

Brian J. McGill

Ecologists have borrowed a powerful tool from physics to calculate how environmental constraints affect the abundance of species.

In the drainage basin of one small river in the center of the North American continent, one can find Kirtland's warbler, which has a total population that seems to fluctuate around a few thousand individuals. In that same area, or indeed almost anywhere else east of the Rocky Mountains up until about 200 years ago, the passenger pigeon thrived with a total population size estimated in the low billions (*1*). This six-orders-of-magnitude discrepancy begs an explanation, especially once we notice that this seems to be one of ecology's few universal laws (see the figure). Every ecosystem in the world, whether at the bottom of the sea or the middle of the Amazonian rain forest, has a few hyperabundant species and many relatively rare species (*2*). Understanding why a species has a particular abundance is the embarrassing and obvious question that ecology cannot yet answer.

The author is in the Department of Biology, McGill University, Montreal, QC H3A 1B1 Canada E-mail: mail@brianmcgill.org

On page 812 of this issue, Shipley *et al.* take a good first step toward an explanation (*3*). The setup is simple. Twelve vineyards were abandoned in southern France over a period ranging from 2 to 42 years ago. These yards slowly returned to natural vegetation, and the relative abundance (percentage of the total plant population $p_1 \ldots p_{30}$) of 30 plant species in these plots was counted. The authors also measured a suite of eight characteristics or traits, such as perennial versus annual, thickness of leaf, and height of plant for each species ($t_{\text{trait,species}}$), for a total of $240 = 8 \times 30$ trait measures. They then calculated the average values of these eight traits ($t_{\text{trait},*}$) for each vineyard as a whole, using the equations:

$$t_{\text{height},*} = p_1 t_{\text{height},1} + \ldots + p_{30} t_{\text{height},30}$$
$$\ldots \quad (1)$$
$$t_{\text{leafthickness},*} = p_1 t_{\text{leafthickness},1} + \ldots + p_{30} t_{\text{leafthickness},30}$$

Next, Shipley *et al.* showed something elegant: These average traits show orderly change over time as the vineyards return to

nature (see their figure 1). They hypothesize, although they do not directly test, that this is due to what they call environmental filtering—for any given environmental context (including field age), there is an "optimal" value for the traits, and species with trait combinations close to this optimum fare better. At this juncture, Shipley *et al*. took a radical departure and did something that ecologists loathe; they borrowed a tool from physics. And not just any tool, but a tool that is still shiny and new in physics called entropy maximization (EM).

Physicists revel in the idea that most of their laws can be reduced to optimization principles (e.g., Lagrange's law of minimum "action" supplants Newton's three laws of motion). It has recently been shown that EM can replace a supercomputer with just a few lines of calculations for modeling how the unequal arrival of solar energy gets redistributed by the atmosphere and oceans (i.e., weather). And this works not only for Earth but for Saturn's moon Titan. This has catapulted EM to prominence in the physical sciences (*4*) after lying on a back shelf for 40 years (*5*). But ecologists who study communities of species tend to regard maximization principles as disreputable (for some good reasons).

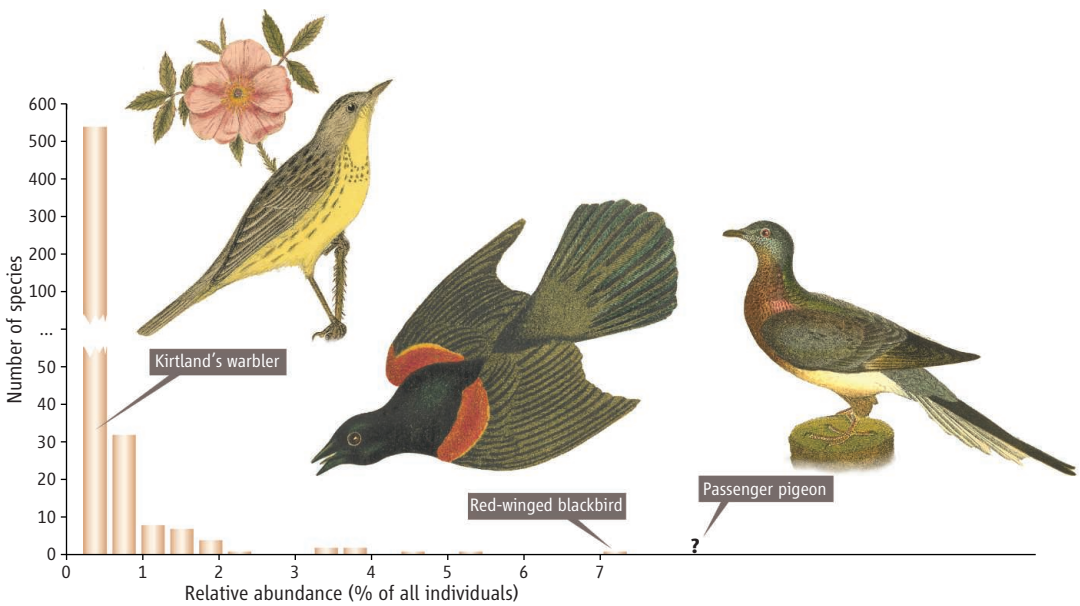
Against this context, Shipley *et al.* boldly apply the EM principle in ecology to predict abundances. EM starts with constraints (what is known about the system) and fills out the rest (our ignorance) by maximizing entropy. The environmental filtering hypothesis used by Shipley *et al*. asserts that traits constrain communities such that abundances are chosen to produce average traits ($t_{\text{trait},*}$) optimal for the given environmental conditions (Eq. 1 in reverse). From high school algebra we know, however, that starting with only eight traits (i.e., equations or constraints) and trying to predict 30 species abundances p_i (i.e., unknowns) is an underdetermined problem; there are an infinite number of solutions arrayed across a 22-dimensional space (22 = 30 – 8). Some additional rule is needed to pick from this infinitude of possible answers. Here is where entropy steps in. Entropy is best thought of as evenness (even distribution of heat across a planet, or even abundances across species, i.e., $p_1 = p_2 = \ldots = p_{30}$). With the mathematical technique known as Lagrange multipliers used to maximize entropy, the predicted relative abundances for all 30 species (equation 5 of Shipley *et al*.) pop right out. The method works so well that Shipley *et al*. can explain 96% of the variation in relative abundances of each species, the kind of result that ecologists usually only lust after.

Why it works is harder to explain. In statistical mechanics entropy has a clear meaning, but in ecology it is a vague concept (despite having been used for years as a measure of evenness). Whether entropy represents species acting randomly and individualistically or communities acting to maximize a collective property such as energy transformation is really just new words in a long-running debate in ecology (*6*).

The report by Shipley *et al*. is exemplary of a more general renaissance occurring right now. The study of abundance had been stuck on three classical approaches: (i) using differential equations to describe the population dynamics of a species has proven good at explaining the variation in abundance of one species over time (that is what differential equations do, after all) but poor at predicting different abundances between species; (ii) finding correlations between traits and abundance has largely failed (*7*), probably because of the focus on one trait at a time, until the work of Shipley *et al*.; and (iii) relying on purely stochastic explanations (each species' abundance is set by a roll of Mother Nature's dice) fails to explain why relative abundances can stay constant for a million years (*6*) or why abundances bounce back almost immediately to formerly high levels after spending a couple of thousand years near extinction while fighting off a pest (*8*).

These classic approaches began to be left behind as the study of abundance was reinvigorated 5 years ago with the introduction of neutral theory (*9, 10*). Neutral theory is an elegant theory that makes strong predictions about abundance but rejects two ideas dear to ecologists: the importance of the environmental context and of species differences (traits). Ecologists have fought back by falsifying neutral theory (*11*) but have not yet put up a fair fight by giving an alternative theory that makes equally strong predictions about abundances while incorporating traits and environment (*12*). Shipley *et al.* just may have made the fight fair.

Other fundamental questions about abundance are finally beginning to be explored as well, such as (i) why abundance varies by about two orders of magnitude across space within a single species, (ii) why abundance of a species changes with temperature, (iii) why abundance and range size are so strongly correlated, (iv) why naturally (not human-caused) rare species such as Kirtland's warbler persist so long, and (v) what factors cause the (always large) portion of rare species to vary by small amounts. This



Wildly different. The bar graph shows a histogram of the relative abundances of more than 500 bird species from the Breeding Bird Survey (*14*), which counts every bird encountered at more than 1400 points across North America. The horizontal axis is relative abundance (percentage of total individuals observed) binned into groups, and the vertical axis indicates the number of species falling into each group of abundances (note broken axis). The left bird is Kirtland's warbler, which falls in the lowest group of abundances. The bottom right bird is the passenger pigeon, which is now extinct but would have been placed off the right end of the graph. The center bird is the red-winged blackbird, which is now the most abundant bird in North America and is 400,000 times more common than the 16 rarest birds in the leftmost bin.

CREDIT: KIRTLAND'S WARBLER, C. J. MAYNARD; PASSENGER PIGEON AND RED-WINGED BLACKBIRD, ALEXANDER WILSON

renewed interest cannot come too soon. Understanding abundance is critical to conservation and global change. It is about time that ecologists start to deliver on our claim that we study "the distribution and abundance of particular species" [(*13*), p. 3].

References

1. *The Birds of North America Online*, A. Poole, Ed. (Cornell Laboratory of Ornithology, Ithaca, 2005) at http://bna.birds.cornell.edu/BNA.
2. M. Tokeshi, *Adv. Ecol. Res.* **24**, 111 (1993).
3. B. Shipley, D. Vile, É. Garnier, *Science* **314**, 812 (2006).
4. R. Lorenz, *Science* **299**, 837 (2003).
5. E. T. Jaynes, in *Statistical Physics,* K. Ford, Ed. (Benjamin, New York, 1963), pp. 181–218.
6. B. J. McGill, E. A. Hadly, B. A. Maurer, *Proc. Nat. Acad. Sci. U.S.A.* **102**, 16701 (2005).
7. B. R. Murray, P. H. Thrall, A. M. Gill, A. B. Nicotra, *Austral Ecol.* **27**, 291 (2002).
8. T. D. Allison, R. E. Moeller, M. B. Davis, *Ecology* **67**, 1101 (1986).
9. G. Bell, *Science* **293**, 2413 (2001).
10. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, 2001).
11. B. J. McGill, B. A. Maurer, M. D. Weiser, *Ecology* **87**, 1411 (2006).
12. B. J. McGill, B. J. Enquist, E. Weiher, M. Westoby, *Trends Ecol. Evol.* **21**, 178 (2006).
13. H. G. Andrewartha, L. C. Birch, *The Ecological Web: More on the Distribution and Abundance of Animals* (Univ. Chicago Press, Chicago, 1984).
14. *Breeding Bird Survey* (Patuxent Wildlife Research Center, Laurel, MD, 2001), available at www.pwrc.usgs.gov/BBS.

10.1126/science.1134920

ASTRONOMY

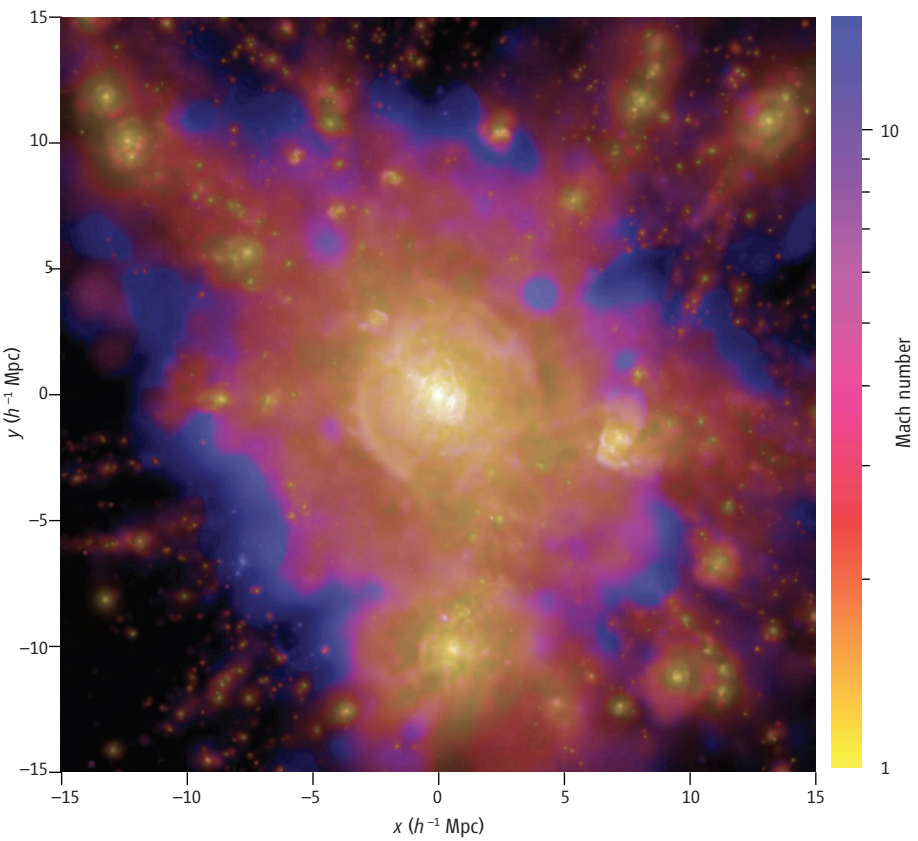
Radio Traces of Cosmic Shock Waves

Colliding and fusing galaxy clusters should produce giant shock waves. The outlines of these waves have now been seen as radio-emitting structures.

Torsten A. Enßlin

The distribution of galaxies in the universe is marked by vast cosmic voids embraced by a network of galaxy filaments and massive galaxy clusters containing up to thousands of galaxies. This inhomogeneous matter distribution emerged from an extremely smooth initial state created by the Big Bang, with relative density fluctuations of only 10^{-5}. This remarkable smoothness was first revealed by the work of the COBE (Cosmic Background Explorer) team, work that was awarded the 2006 Nobel prize in physics. Over billions of years, the initially tiny density variations grew drastically through gravitational attraction of neighboring matter. Larger and larger structures still form today as a result of the violent merging of galaxies and clusters of galaxies. In addition, there is a continuous accretion flow of gas falling onto galaxy clusters out of the dilute intergalactic medium. On page 791 of this issue, Bagchi *et al*. (*1*) report the detection of giant radio structures around a galaxy cluster that probably trace shock waves caused by such energetic collisions, mergers, and movement of gas.

Gas falling into the gravitational wells of galaxy clusters can reach velocities of up to a few thousand kilometers per second. When it collides with the hot and ionized gas at a temperature of 10^7 to 10^8 K within clusters, shock waves form and heat the infalling gas to similar temperatures. Magnetic fields in the gas may permit a small fraction of the thermal gas particles to scatter back into the upstream region of the shock wave and to undergo the energizing shock compression again and again. This so-called diffusive shock acceleration process produces nonthermal particles with an energy spectrum easily extending to ultrarelativistic energies, where particle energies exceed their rest mass energies by large factors. Although the number of these relativistic particles is small compared with the thermal ones, they can account for a substantial fraction of the dissipated shock energy.



Energetic events. Energy dissipation by cosmic shock waves around a massive galaxy cluster and two smaller infalling systems in a numerical simulation. The brightness scales logarithmically with the dissipation rate, the colors indicate the (dissipation weighted) shock Mach numbers. Although most of the energy dissipation occurs within a few megaparsecs around the cluster centers, the surrounding accretion shock waves have the highest Mach numbers (blue structures). A pair of merger-induced shock waves can also be seen roughly 3 Mpc away from the center of the main cluster.

CREDIT: C. PFROMMER/CANADIAN INSTITUTE FOR THEORETICAL ASTROPHYSICS

The author is at Max-Planck-Institut für Astrophysik, Garching, Germany. E-mail: ensslin@mpa-gardching.mpg.de

The efficiency of diffusive shock acceleration increases rapidly with the ratio of the shock velocity to the initial sound speed, a quantity known as the Mach number. Although most of the energy of the cosmic structure formation is dissipated in the centers of galaxy clusters, the shock waves in the outskirts and especially the accretion shocks have much higher Mach numbers and therefore should be more efficient particle accelerators, as can be seen in the figure (*2*).

Electrons, which can be accelerated to energies of 10^4 to 10^5 times their rest mass, produce radio emission due to their gyromotion in intergalactic magnetic fields. Such radio emission in galaxy clusters has been observed since the 1970s (*3*) and named cluster radio relics. However, only recently has the association with cluster merger shock waves been recognized (*4*).

Bagchi *et al.* have found a pair of giant radio structures and propose that the double relic in galaxy Abell 3376 may be emission from the accretion shock of the cluster. This dual radio morphology may be caused by the stronger matter flow onto the cluster along an embedding galaxy filament. If this interpretation is correct, it would be a remarkable finding, because it would imply the presence of magnetic fields in the infalling gas, whereas magnetic fields have so far only been detected within galaxy clusters. Furthermore, we would have the first observational identification of an accretion shock wave. Accretion shock waves are very interesting because they may be the origin of the still-mysterious ultra-high-energy cosmic rays (*5*), which are protons with energies up to 10^{20} eV. The highest energy electrons from such shocks can scatter photons of the cosmic microwave background into gamma-ray bands and thereby contribute to the observed and still unresolved gamma-ray background (*6*, *7*). As a result, the radio relics in Abell 3376 mark locations to be monitored in the future for all kinds of high-energy radiation.

There is another plausible explanation for the double relics, however. In the late stage of a violent merger of similarly sized galaxy clusters, an outgoing pair of shock waves emerges. These shock waves steepen as they run into the more dilute gas of the cluster outskirts, similar to tsunami waves propagating into shallower water. A resulting pair of radio relics was indeed observed in a morphologically similar merging cluster, Abell 3667 (*8*), and well reproduced by numerical simulations (*9*). Possibly, the relics in Abell 3376 are also of this type.

In any case, it is exciting that the radio relics in Abell 3376 provide us with direct insight into the fluid dynamics of cosmic structure formation. This important and surprising observation gives a foretaste of the radio glow of the cosmic large-scale structure (*10*), which one hopes to discern with the next generation radio telescopes such as the Low Frequency Array [LOFAR (*11*)], the Long Wavelength Array [LWA (*12*)], and the Square Kilometre Array [SKA (*13*)].

References

1. J. Bagchi, F. Durret, G. B. Lima Neto, S. Paul, *Science* **314**, 791 (2006).
2. C. Pfrommer, V. Springel, T. A. Enßlin, M. Jubelgas, *Mon. Not. R. Astron. Soc.* **367**, 113 (2006).
3. C. H. Costain, A. H. Bridle, P. Feldman, *Astrophys. J.* **175**, L15 (1972).
4. T. A. Enßlin, P. L. Biermann, U. Klein, S. Kohle, *Astron. Astrophys.* **332**, 395 (1998).
5. H. Kang, J. P. Rachen, P. L. Biermann, *Mon. Not. R. Astron. Soc.* **286**, 257 (1997).
6. A. Loeb, E. Waxman, *Nature* **405**, 156 (2000).
7. F. Miniati, *Mon. Not. R. Astron. Soc.* **337**, 199 (2002).
8. R. T. Schilizzi, W. B. McAdam, *R. Astron. Soc., Memoirs* **79**, 1 (1975).
9. K. Roettiger, J. O. Burns, J. M. Stone, *Astrophys. J.* **518**, 603 (1999).
10. E. Waxman, A. Loeb, *Astrophys. J.* **545**, L11 (2000).
11. www.lofar.org
12. http://lwa.unm.edu
13. www.skatelescope.org

10.1126/science.1133949

BIOMEDICINE

Life, the Universe, and Body Temperature

A hypothermic life-style may lead to a longer life. How good are the prospects?

Clifford B. Saper

In his book *Life, the Universe, and Everything*, Douglas Adams describes an advanced civilization that asks a supercomputer to calculate an answer to the Ultimate Question of "life, the universe, and everything." After several million years of calculation, the computer answers: "42."

Enhanced online at www.sciencemag.org/cgi/content/full/314/5800/773

A similarly inscrutable constant that we face in everyday life is 37, the mean body temperature measured in degrees Celsius of humans and most other mammals. We tend to take this number for granted, as it is always in the same, narrow range, until, of course, we become ill with a fever. We then take medications, usually inhibitors of prostaglandin synthesis (aspirin, ibuprofen, etc.), which typically brings our body temperature back to normal. But why is 37°C "normal", and is this truly the optimal operating temperature for our bodies?

On page 825 of this issue, Conti *et al.* (*1*) question this dogma. Surprisingly, their results suggest that our usual body temperature may not be optimal, at least in determining our life span. Their work is based on a growing revolution in our understanding of how the brain controls body temperature. Although it has been known for decades that the preoptic area—the most rostral tip of the hypothalamus—is both thermosensitive and necessary for maintaining normal body temperature, the details of the neural circuits that control body temperature have only recently begun to be elucidated (*2*). It is now understood that neurons in the medial preoptic region have an intense inhibitory effect on thermogenic responses (see the figure). Other neurons in the middle part of the hypothalamus, including the paraventricular and dorsomedial nuclei, have an excitatory effect on thermogenesis, but are normally held in check by the preoptic neurons. The interplay between the thermogenic neurons and those in the medial preoptic nucleus that hold them in check is critical in controlling body temperature under a wide range of conditions. The hypothalamic sites, furthermore, have descending inputs to brainstem and spinal areas that control autonomic thermoregulatory responses. By shifting blood flow to cutaneous blood vessels, heat can be exhausted,

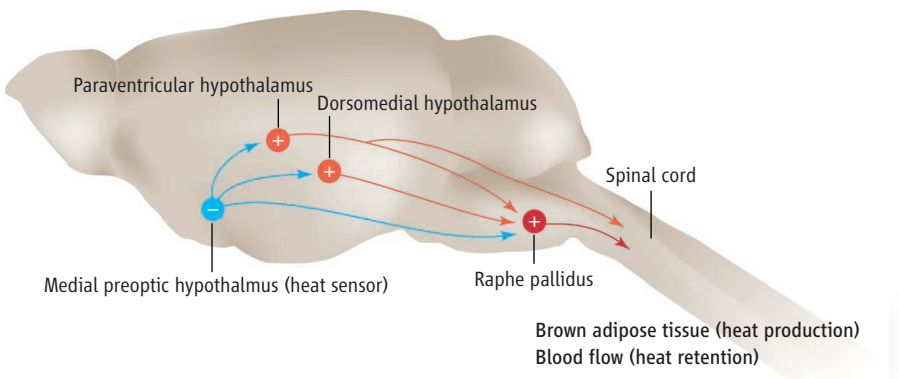
The author is in the Department of Neurology, Division of Sleep Medicine, and Program in Neuroscience, Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA. E-mail: csaper@bidmc.harvard.edu

whereas heat retention is promoted by shifting blood flow to deep blood vessels (hence fingers and toes turn blue in the cold).

Thermogenesis is subserved by neural inputs to brown adipose tissue, at least in small mammals, where β_3 adrenergic receptors mediate production of uncoupling protein 1 (UCP-1). UCP-1 allows mitochondria in brown adipose tissue to convert adenosine 5′-triphosphate (ATP) to heat, rather than to energy for performing work. Thus, small mammals that lack sufficient mass for heat retention carry portable heaters in the form of brown adipose tissue that allow them to avoid hypothermia.

Here is where the intervention engineered by Conti *et al*. comes in. They produced transgenic mice in which expression of the UCP-2 gene (closely related to UCP-1) is placed under the control of the promoter for hypocretins (also called orexins). Hypocretins are peptides that are produced only by cells in the lateral hypothalamus (*3*). By placing UCP-2 expression under the control of this promoter, the investigators effectively placed a small heater into the hypothalamus. As their data show, this caused heating of the preoptic area, a region in which previous work had shown that insertion of heat probes would cause a reduction in body temperature. The result is that the transgenic animals expressing the UCP-2 gene had a continuous reduction in body temperature by 0.3° to 0.5°C.

Surprisingly, there has been little previous work on the effects of life-long hypothermia on other physiological functions in mammals, mainly because the brain normally maintains a constant body temperature so thoroughly that any change from this condition is rather difficult to achieve. In the Conti *et al*. experiments, the hypothermic transgenic mice showed no change in food intake or physical activity, but their body weight did increase, presumably due to a lower basal metabolic rate. One might expect, given the accumulated evidence that increased weight correlates with a variety of disorders that shorten life (*4*), that the hypothermic mice might have had a shorter life span. But in fact, the opposite was the case. Not only did the hypothermic animals live about 3 months longer than the 27-month mean life span for control mice, but they also had a parallel mortality rate, as if the mortality curve had been shifted to the right by about 10% of the life span. A corresponding percentage increase in human life span would be 7 to 8 years, a sizable change in longevity. If this were due to a delay in senescence, as the shift of the mortality curve in the hypothermic mice would suggest, rather than merely a prolongation of it, this could substantially increase lifetime productivity in humans.



Temperature control center. In the mammalian brain (mouse brain shown), a series of neural pathways (red) control the body's autonomic responses that regulate heat conservation and production, respectively. Cells in the paraventricular and dorsomedial hypothalamic nuclei, and in the raphe pallidus, signal to sympathetic preganglionic neurons in the spinal cord to control thermogenesis. These pathways are in turn regulated by an inhibitory input (blue) from the medial preoptic hypothalamus that is responsive to preoptic temperature.

Perhaps we should not be surprised by this result, because low body temperature does prolong life span in poikilothermic fish (in which body temperature fluctuates with that of the external environment) (*5*). Homeotherms with a restricted caloric intake develop a low body temperature and also have a prolonged life span (*6*). But few people would choose a life-style that limits their caloric intake, and Conti *et al*. provide the first test of the hypothesis that lowering the body temperature of a mammal prolongs life. The substantial increase of life span raises the question of whether mild hypothermia of 0.3° to 0.5°C might be easier to tolerate than a lifetime of starvation, as a way to increase longevity. Although at present there is no practical way for humans to achieve prolonged hypothermia, the results of Conti *et al*. suggest a potential gene therapy approach. One could imagine, for example, stereotaxic injections into the hypothalamus of an adeno-associated virus or lentivirus engineered to provide long-term expression of an uncoupling protein, to warm the hypothalamus just enough to extend life span.

If life-span extension could be this simple, one might wonder whether 37°C is indeed the optimal body temperature for humans, and why evolution has not selected for a lower body temperature and longer life span. However, there would be little evolutionary pressure to extend the number of years of life after reproduction is finished. A more important question for humans contemplating a hypothermic life-style might be whether the lower body temperature in the UCP-2 transgenic mice might cause other physiological or behavioral problems, such as changes in reproductive physiology, which might select against a lower body temperature. The reasons for the remarkable stability of body temperature among mammals, and why this temperature has been selected by evolution, remain obscure, although one would certainly want to know the consequences of hypothermia before pursuing it as a way to increase life span.

In Adams's book, the scientists ask the supercomputer how the answer to the Ultimate Question of "life, the universe, and everything" could be "42." The computer answers that to understand the answer, they have to know what the Ultimate Question is, and that it will take several million more years to determine that. We hope that we will not have to wait as long to understand why evolution has designed us with a body temperature of 37°C. The new and unexpected vista on the relation between body temperature and longevity opened up by the report of Conti *et al*. may help expedite the process. This work also holds out the tantalizing promise that we may be able to achieve a longer life span, if we were only to be a little cooler about it.

References

1. B. Conti *et al*., *Science* **314**, 825 (2006).
2. S. F. Morrison, *News Physiol. Sci.* **19**, 67 (2004).
3. J. T. Willie, R. M. Chemelli, C. M. Sinton, M. Yanagisawa, *Annu. Rev. Neurosci.* **24**, 429 (2001).
4. G. A. Bray, T. Bellanger, *Endocrine* **29**, 109 (2006).
5. R. K. Liu, R. L. Walford, *J. Gerontol.* **30**, 129 (1975).
6. B. A. Rikke *et al*., *Mech. Ageing Dev.* **124**, 663 (2003).

10.1126/science.1135375

A Century of Alzheimer's Disease

Michel Goedert[1]* and Maria Grazia Spillantini[2]

One hundred years ago a small group of psychiatrists described the abnormal protein deposits in the brain that define the most common neurodegenerative diseases. Over the past 25 years, it has become clear that the proteins forming the deposits are central to the disease process. Amyloid-β and tau make up the plaques and tangles of Alzheimer's disease, where these normally soluble proteins assemble into amyloid-like filaments. Tau inclusions are also found in a number of related disorders. Genetic studies have shown that dysfunction of amyloid-β or tau is sufficient to cause dementia. The ongoing molecular dissection of the neurodegenerative pathways is expected to lead to a true understanding of disease pathogenesis.

On 3 November 1906, at the 37th meeting of the Society of Southwest German Psychiatrists in Tübingen, Germany, Alois Alzheimer presented the clinical and neuropathological characteristics of the disease (*1*, *2*) that Emil Kraepelin subsequently named after him (*3*). Alzheimer's disease (AD) is now the most common neurodegenerative disease, with more than 20 million cases worldwide. At the time of his lecture, Alzheimer was head of the Anatomical Laboratory at the Royal Psychiatric Clinic of the University of Munich. He had moved there in 1903 after having spent 14 years at the Municipal Institution for the Mentally Ill and Epileptics in Frankfurt, where Franz Nissl had introduced him to brain histopathology. In November 1901, Alzheimer admitted Auguste D., a 51-year-old patient, to the Frankfurt hospital because of progressive memory loss, focal symptoms, delusions, and hallucinations. After the death of Auguste D. in April 1906, her brain was sent to Munich for analysis. Alzheimer's use of the silver staining method developed by Max Bielschowsky 4 years earlier (*4*) was crucial for the identification of neuritic plaques and neurofibrillary tangles, the defining neuropathological characteristics of the disease. Whereas plaques had been reported before, first by Blocq and Marinesco in an elderly patient with epilepsy (*5*), Alzheimer was the first to describe the tangle pathology. In 1911, he found a different type of nerve cell inclusion in two cases with focal degeneration of the cerebral cortex (*2*). This is now known as the Pick body (even though it was first described by Alzheimer) and the clinicopathological entity is known as Pick's disease, after Arnold Pick, who first described it in 1892 (*6*). Pick's disease belongs to the spectrum of frontotemporal lobar degeneration (FTLD).

The presence of abnormal deposits helped greatly with disease classification (*7*). However, their molecular composition and role in the pathological process remained unknown. Over the past 25 years, a basic understanding has emerged from the coming together of two independent lines of research. First, the molecular study of the deposits led to the identification of their major components. Second, the study of rare, inherited forms of disease resulted in the discovery of the causative gene defects. In most cases, the defective genes encode the major components of the pathological lesions or factors that change their levels. It follows that a toxic property of the proteins that make up the filamentous lesions underlies the inherited disease cases. A similar toxic property may also cause the much more common sporadic forms of disease. Here we review the evidence implicating amyloid-β and tau in neurodegeneration.

Abnormal Filaments

In the electron microscope, plaques and tangles contain abnormal filaments (*8*, *9*). Plaque filaments are extracellular and have the molecular fine structure of amyloid. This term refers to filaments with a diameter of around 10 nm that have a cross-β structure and characteristic dye-binding properties. Most tangle filaments have a paired helical morphology and are also amyloid-like. Paired helical filaments are present in nerve cell bodies, as well as in neurites in the neuropil and at the periphery of neuritic plaques. After the identification of filaments (*8*, *9*), it took another 20 years before their major components were known. The identification of amyloid-β as the major plaque component and tau as the major tangle component ushered in the modern era of research on AD (Fig. 1A). Filamentous tau deposits are also present in a number of other neurodegenerative disorders, including Pick's disease (Fig. 1B).

Amyloid-β

Amyloid-β is 40 to 42 amino acids in length and is generated by proteolytic cleavage of the much larger amyloid precursor protein (APP), a transmembrane protein of unknown function with a single membrane-spanning domain (*10*–*13*) (Fig. 2A). The N terminus of amyloid-β is located in the extracellular domain of APP, 28 amino acids from the transmembrane region, and its C terminus is in the transmembrane region. The enzymes whose activity gives rise to the N and C termini are called β-secretase and γ-secretase, respectively. A third enzyme, α-secretase, cleaves between residues 16 and 17, precluding amyloid-β formation. The major species of amyloid-β are 40 or 42 amino acids long, and it is the more amyloidogenic 42–amino acid form (with its two additional hydrophobic amino acids) that is deposited first (*14*). In the three-dimensional structure of the amyloid-β fibril, residues 1 to 17 are disordered, with residues 18 to 42 forming a β-strand–turn–β-strand motif that contains two parallel β sheets formed by residues 18 to 26 and 31 to 42 (*15*).

Mapping of the *APP* gene to chromosome 21, together with observation of plaques and tangles in most elderly individuals with Down's syndrome (trisomy of chromosome 21), suggested an important role for amyloid-β. However, direct genetic evidence was lacking. It came from work on hereditary cerebral hemorrhage with amyloidosis–Dutch type (HCHWA-D), a rare condition characterized by recurrent hemorrhages resulting from the deposition of amyloid-β in cerebral blood vessel walls. HCHWA-D is caused by a missense mutation in the amyloid-β portion of APP (*16*). Six years after the purification of amyloid-β from meningeal blood vessels of AD brains (*10*), this was the second time that cerebral blood vessels were found to play a crucial role in advancing the understanding of AD. Although HCHWA-D is characterized by amyloid-β deposits in the walls of cerebral microvessels, it differs from AD in several respects. Thus, when present, dementia is vascular in origin. Furthermore, plaques are sparse and tangles absent.

In the late 1980s, it was speculated that mutations in the *APP* gene would also be found in familial AD, some cases of which had been linked to chromosome 21 (*17*, *18*). The first such mutations were soon identified (*19*–*21*) (Fig. 2, B and C). Amyloid-β is a normal, secreted product (*22*–*24*), which suggests that it has a (still unknown) physiological function. *APP* mutations increase amyloid-β production or lead to an increased proportion of amyloid-β ending at residue 42 (*25*, *26*). Most mutations flank the amyloid-β region, with the secreted peptide being the wild type. However, several mutations are within amyloid-β itself. Like the HCHWA-D mutation, some of these mutations have little effect on APP processing but increase the propensity of amyloid-β to form fibrils (*27*). Missense mutations in amyloid-β lead to vascular deposits, parenchymal plaques, or both. Twenty missense mutations in the *APP* gene have been described (Fig. 2C). Recently, increased gene dosage was identified as another cause of disease (Fig. 2B). Duplication of the *APP* gene gives rise to amyloid-β deposition in brain neuropil, cerebral blood vessels, or both locations, with clinical pictures of AD or recurrent brain hemorrhages (*28*, *29*). These findings are remi-

[1]Laboratory of Molecular Biology, Medical Research Council, Cambridge CB2 2QH, UK. [2]Cambridge Centre for Brain Repair, Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 2PY, UK.

*To whom correspondence should be addressed. E-mail: mg@mrc-lmb.cam.ac.uk

niscent of Down's syndrome, although brain hemorrhages are only rarely observed. They underscore the need to understand more about the factors that determine whether amyloid-β is deposited in brain or vasculature. Neuronally derived amyloid-β is transported to the vasculature, where it is cleared via transport into the blood or via the perivascular fluid drainage pathway (*30*). These findings have been replicated to some extent in transgenic mice. Expression of mutant human APP in nerve cells leads to amyloid plaque and blood vessel wall deposits (*31*, *32*). However, tangles and extensive nerve cell loss have not been observed in these mouse lines.

Mutations in the *APP* gene account for only a minority of familial AD cases. Linkage studies established the presence of a major disease locus on chromosome 14 (*33*), and positional cloning led to the identification of mutations in the *presenilin-1* gene, which encodes a polytopic membrane protein (*34*). Mutations in *presenilin-1* are the most common cause of familial AD. Mutations in the related *presenilin-2* gene also give rise to AD (*35*, *36*). More than 160 mutations in the *presenilin* genes have been identified. Presenilins are central components of the atypical aspartyl protease complexes responsible for the γ-secretase cleavage of APP (*37*, *38*), but other transmembrane proteins are also γ-secretase substrates. *Presenilin* gene mutations increase the ratio of amyloid-β 42 to amyloid-β 40, and this appears to result from a change in function (*39*) that manifests itself in reduced γ-secretase activity. In preclinical cases with *presenilin-1* mutations, deposition of amyloid-β 42 is an early event (*40*, *41*). The phenotypic spectrum associated with *presenilin* gene mutations may extend beyond AD to encompass cases of FTLD with tau deposits (*42*). If confirmed, this would indicate that these mutations can cause disease through amyloid-β–independent effects. Support for this notion comes from transgenic animal models, which have suggested that a reduction in γ-secretase activity can lead to the hyperphosphorylation of tau in the absence of amyloid-β deposits (*43*). Unlike the presenilins, no disease-causing mutations have been identified in the aspartyl protease BACE1, which is identical with β-secretase (*44*).

Taken as a whole, the work on familial AD forms the bedrock of the amyloid cascade hypothesis (*45*), which holds that an increase in amyloid-β 42 triggers all cases of AD, with tangle formation, nerve cell degeneration, and dementia being downstream events.

Tau

It took several years of work before it was clear that the paired helical filaments are made of full-length, hyperphosphorylated tau, a protein involved in microtubule assembly and stabilization (*46*–*53*). In the human brain, six tau isoforms are produced from a single gene through alternative mRNA splicing (*54*) (Fig. 3A). They fall into two groups on the basis of numbers of microtubule-binding repeats, with three isoforms having three repeats each and three isoforms having four repeats each. The presence or absence of N-terminal inserts distinguishes the three isoforms in each group. In the normal human brain, similar levels of three- and four-repeat isoforms are expressed. In tau filaments from AD brains, all six isoforms are present in proportions similar to those in normal brains. Filamentous tau deposits are also found in a number of other neurodegenerative diseases, including progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), Pick's disease, argyrophilic grain disease (AGD), and the Parkinson-dementia complex of Guam (*55*). In PSP, CBD, and AGD, the deposits are present in nerve cells and glial cells, whereas in AD, Pick's disease, and the Parkinsonism-dementia complex of Guam they are largely confined to nerve cells. Unlike AD, these diseases all lack amyloid-β pathology. Besides AD, several other neurodegenerative diseases are associated with extracellular protein deposits, such as the Abri peptide in familial British dementia and the prion protein (PrP) in Gerstmann-Sträussler-Scheinker (GSS) disease. As in AD, abundant deposits of tau form in nerve cell bodies and around plaques in familial British dementia and in GSS disease caused by certain *PrP* gene mutations (*56*, *57*).

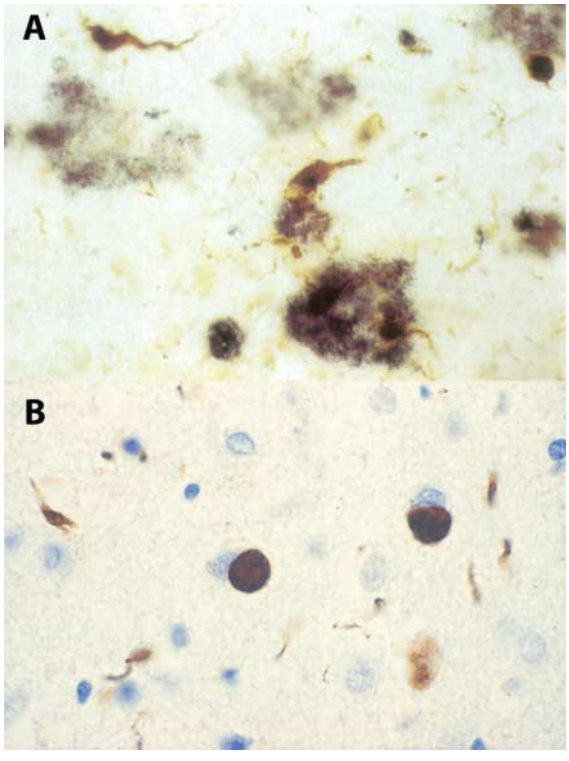


Fig. 1. The abnormal deposits that Alzheimer described. (**A**) Neuritic plaques made of amyloid-β (blue) and neurofibrillary tangles made of tau (brown) in Alzheimer's disease. (**B**) Pick bodies and neurites made of tau (brown) in Pick's disease.

Hyperphosphorylation of tau is common to all diseases with tau filaments and may be required for toxicity. Much is known about phosphorylation sites and candidate protein kinases and phosphatases, auguring well for the development of preventive strategies aimed at reducing tau phosphorylation (*58*). Whereas the phosphorylated sites in tau are similar in the different diseases, the isoform composition of tau filaments differs. In PSP, CBD, and AGD, four-repeat tau isoforms are present, whereas tau isoforms with three repeats are found in Pick's disease. All six isoforms are present in Parkinsonism-dementia complex of Guam, familial British dementia, and cases of GSS disease with tau deposits. The molecular dissection of tau filaments gave a complete description of their composition and provided clues about the mechanisms underlying their formation. However, the relevance of tau dysfunction for the etiology and pathogenesis of AD and related disorders had remained unclear. Such a connection had been suspected because the distribution and abundance of tau pathology correlated well with nerve cell degeneration and clinical symptoms (*59*). However, the identification of mutations in the genes encoding APP and presenilin, and the presence of tau deposits in a number of apparently unrelated disorders, cast doubt on the importance of tau.

The finding that mutations in the *Tau* gene cause the inherited frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17) removed this doubt (*60*–*62*). To date, 39 such mutations have been described (Fig. 3B). FTDP-17, which belongs to the FTLD spectrum of diseases, is quite varied. It can present predominantly as a dementing disorder, a parkinsonian disease, or a condition with motor neuron disease–like symptoms. Neurological syndromes similar to PSP, CBD, and Pick's disease have also been described. Filamentous tau inclusions are invariably present in the absence of amyloid-β deposits. Depending on the mutations, the inclusions are present in nerve cells or in nerve cells and glia, and consist of three-repeat tau, four-repeat tau, or all six isoforms. The different isoform compositions are reflected in varied filament morphologies. *Tau* mutations are located in the coding region or the intron flanking alternatively spliced exon 10. The latter encodes the microtubule-binding repeat included in four-repeat tau. Functionally, mutations fall into two largely nonoverlapping categories: those that influence the alternative splicing of tau pre-mRNA, and those whose

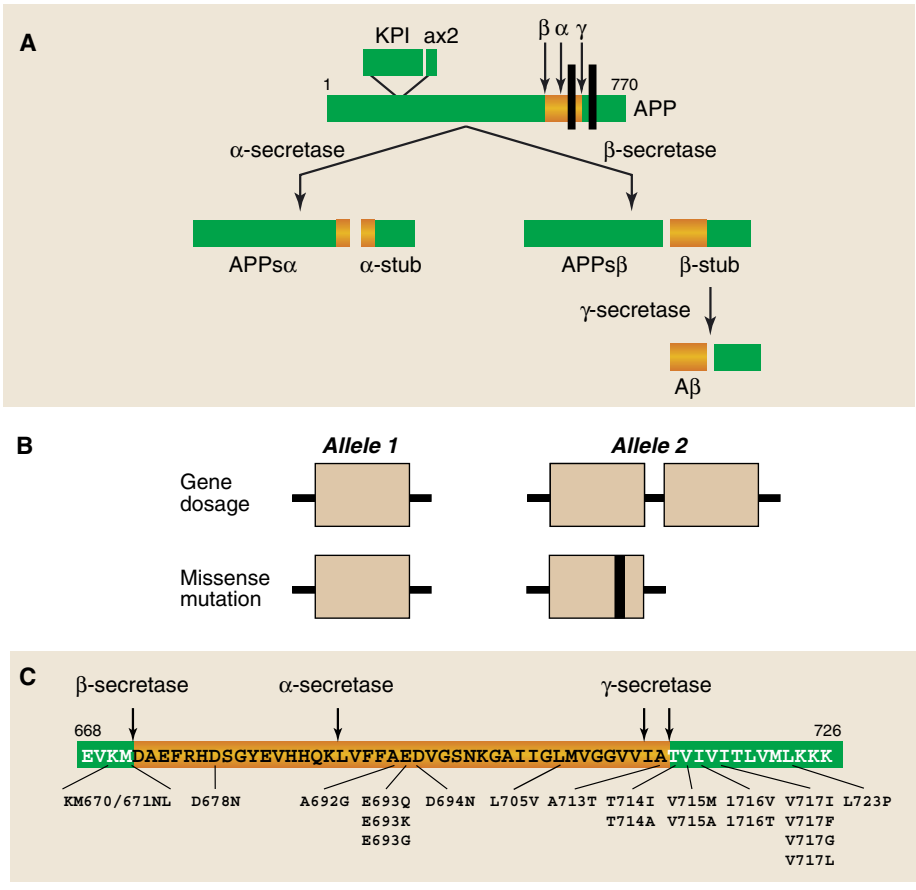


Fig. 2. Amyloid-β. (**A**) Generation of amyloid-β (Aβ) from the amyloid precursor protein (APP). Cleavage by β-secretase generates the N terminus and intramembranous cleavage by γ-secretase gives rise to the C terminus of amyloid-β. Cleavage by α-secretase precludes Aβ formation. (**B**) Duplication of the *APP* gene and missense mutations (black box) in the *APP* gene cause inherited forms of Alzheimer's disease and cerebral amyloid angiopathy. (**C**) Twenty missense mutations in *APP* are shown. Single-letter abbreviations for amino acid residues: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

primary effect is at the protein level. In accordance with their location in the microtubule-binding region, most missense mutations reduce the ability of tau to interact with microtubules (*63*). Some mutations also promote aggregation into filaments. Intronic mutations and most coding region mutations in exon 10 increase the splicing of exon 10, leading to the relative overproduction of four-repeat tau (*61*, *62*, *64*). In the normal brain, a correct ratio of three-repeat to four-repeat tau isoforms is essential for preventing neurodegeneration and dementia. Multiplications of *Tau* have so far not been reported. Although the pathway leading from a mutation in *Tau* to neurodegeneration is only incompletely known, it appears likely that a reduced ability to interact with microtubules is necessary for setting in motion the gain of toxic function that will cause neurodegeneration. This work is relevant beyond FTDP-17, because it shows that whenever filamentous tau inclusions form in the brain, abnormalities in tau are directly involved in the ensuing neurodegeneration.

FTLD itself is genetically heterogeneous, with a substantial number of cases exhibiting tau-negative, ubiquitin-positive nerve cell inclusions. Mutations in the genes encoding the apparently unconnected p97 (*65*), CHMP2B (charged multivesicular body protein 2B) (*66*), and, in particular, progranulin (*67*, *68*) cause these forms of FTLD. In contrast to *Tau* mutations, they all appear to lead to disease through loss of function of the mutant allele.

Haplotypes H1 and H2 characterize the *Tau* gene in populations of European descent (*69*). They are the result of a 900-kb genomic inversion polymorphism that encompasses *Tau* (*70*). Heterozygous microdeletions in this region give rise to a form of mental retardation (*71*–*73*). These findings point to a possible role for tau in brain development and are consistent with the notion that FTDP-17 is caused by a gain of toxic function of tau. Inheritance of H1 is a risk factor for PSP and CBD (*69*, *74*, *75*). An association has also been described between H1 and idiopathic Parkinson's disease (*76*), a disease without tau pathology. H1 has been shown to be more effective than H2 at driving the expression of a reporter gene, which suggests that higher levels of tau are expressed from H1 (*77*). However, it remains unclear how this could explain the preferential deposition of four-repeat tau in PSP and CBD.

The work on FTDP-17 has led to the development of robust transgenic mouse models that replicate the essential molecular and cellular features of the human tauopathies, including tau hyperphosphorylation, filament formation, and extensive nerve cell loss (*78*, *79*). The crossing of lines expressing mutant tau with lines expressing mutant APP results in enhanced tau pathology (*80*).

Sporadic Alzheimer's Disease

Most cases of AD are sporadic, with dominantly inherited forms accounting for less than 1% of the total. Inheritance of the ε4 allele of apolipoprotein E (*APOE*) is the only well-established genetic risk factor for sporadic AD (*81*), but its mode of action is unknown. Amyloid-β deposits are more abundant in ε4-positive than in ε4-negative cases (*82*). In addition, apoE4 is associated with a number of other factors that may contribute to AD pathology, including low glucose usage, mitochondrial abnormalities, and cytoskeletal dysfunction (*83*).

Age is a major risk factor for AD, and small numbers of plaques and tangles form in most individuals as they grow older (*59*). Tau pathology appears first in the transentorhinal region, from where it spreads to the hippocampus and amygdala, followed by neocortical areas. Amyloid-β deposits tend to appear first in the neocortex. Both types of inclusion seem to form independently, with tangles appearing first. At later stages, extensive amyloid-β deposition in the neocortex has been reported to precede severe tangle pathology (*84*), leading to the suggestion that amyloid-β deposition may exacerbate age-related tau pathology. This would be consistent with what is known from cases with *APP* gene mutations and duplications, where overproduction of amyloid-β 42 is upstream of tau dysfunction. Mutations in *Tau*, on the other hand, lead to filament formation, neurodegeneration, and dementia but do not give rise to amyloid-β deposits. An outstanding question relates to the molecular nature of the neurotoxic species. In recent years, evidence has accumulated that suggests that oligomeric species of amyloid-β and tau may be major culprits (*85*). For tau, it appears likely that the mere presence of abnormal filaments in nerve cell processes is also detrimental to the parent cell, if only because they are space-occupying lesions that are bound to interfere with axonal transport.

In AD, neurodegeneration is estimated to start 20 to 30 years before the appearance of the first clinical symptoms. The early clinical phase

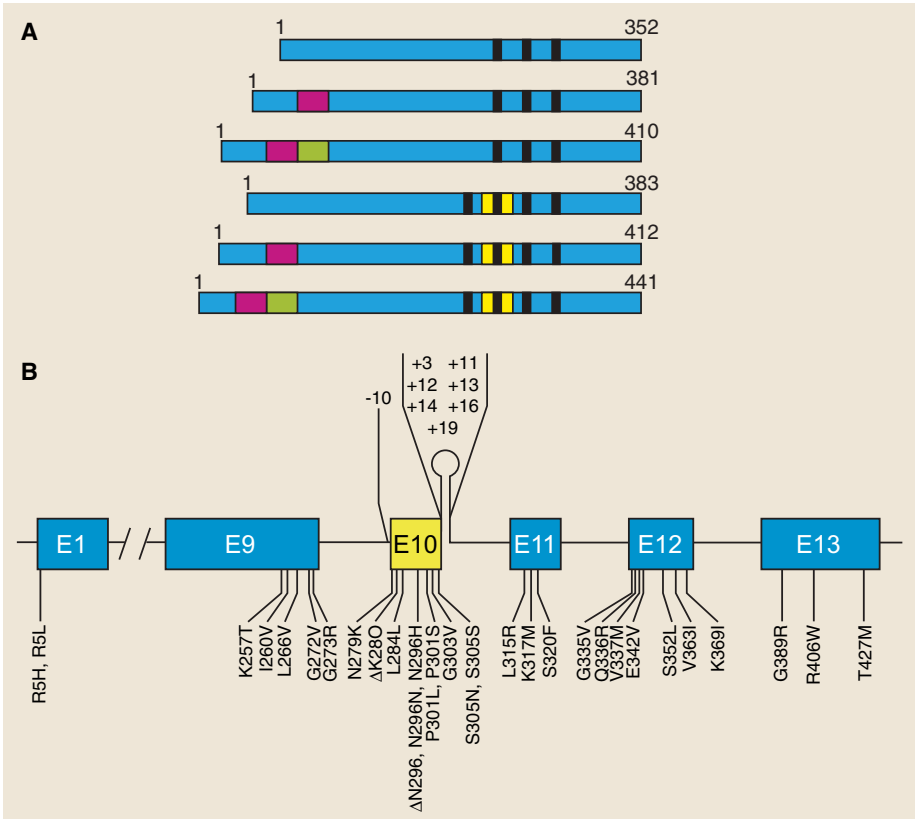


Fig. 3. Tau. (**A**) The six tau isoforms expressed in adult human brain. Alternatively spliced exons are shown in red, green, and yellow, respectively, and the microtubule-binding repeats are indicated by black bars. (**B**) Mutations in the *Tau* gene in frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17). Thirty-one coding region mutations in exons (E) 1, 9, 10, 11, 12, and 13 and eight intronic mutations flanking E10 are shown.

is often called amnestic mild cognitive impairment (aMCI) (*86*). The neuropathological features of aMCI are intermediate between those of normal aging and AD, in that tau deposits are abundant in the entorhinal cortex and hippocampus and some amyloid-β deposits are present in the neocortex (*87*). For aMCI, the regional distribution of tau deposits correlates better with the degree of cognitive impairment than does the amyloid-β load. It has been suggested that the transition from aMCI to AD occurs when tau pathology spreads beyond the medial temporal lobe. Work has so far concentrated on the presence of deposits. In the future, it will be important to measure levels of amyloid-β and tau oligomers in aMCI.

The long presymptomatic phase of AD augurs well for the development of preventive strategies. To test their effectiveness, it will be necessary to identify neuropathological abnormalities before the development of cognitive changes. Use of Pittsburgh compound B (PIB), a thioflavin T derivative, has already resulted in the visualization of amyloid-β deposits in patients with AD and in some nondemented elderly individuals (*88*), which suggests that imaging with PIB can detect clinical and preclinical disease. In the future, it may also become possible to image tau deposits in the living human brain.

Closing Remarks

The protein deposits described by Alzheimer are at the center of current work. Although much has been learned, major questions remain. Perhaps the greatest unknown relates to the links between amyloid-β and tau. Another important question concerns the mechanisms that determine the selective vulnerability of defined neuronal and glial populations. A related issue has to do with the molecular species that cause nerve cell degeneration. During his lifetime, Alzheimer was best known for his clinicopathological studies of neurosyphilis, then a pressing problem in psychiatry and the prime example of an organic brain disorder. It receded after advances in microbiology and the advent of chemotherapeutics and antibiotics. The hope is that in the not too distant future, on the basis of the knowledge gained, safe and effective treatments will also become available for AD and related disorders.

References and Notes

1. A. Alzheimer, *Allg. Z. Psychiatr.* **64**, 146 (1907).
2. A. Alzheimer, *Z. Ges. Neurol. Psychiat.* **4**, 356 (1911).
3. E. Kraepelin, *Psychiatrie. Ein Lehrbuch für Studierende und Ärzte. II. Band* (Barth Verlag, Leipzig, 1910).
4. M. Bielschowsky, *Neurol. Centralbl.* **21**, 579 (1902).
5. P. Blocq, G. Marinesco, *Sem. Méd.* **12**, 445 (1892).
6. A. Pick, *Prager Med. Wochenschr.* **17**, 165 (1892).
7. G. Blessed, B. E. Tomlinson, M. Roth, *Br. J. Psychiatry* **114**, 797 (1968).
8. M. Kidd, *Nature* **197**, 192 (1963).
9. R. D. Terry, N. K. Gonatas, M. Weiss, *Am. J. Pathol.* **44**, 269 (1964).
10. G. G. Glenner, C. W. Wong, *Biochem. Biophys. Res. Commun.* **120**, 885 (1984).
11. G. G. Glenner, C. W. Wong, *Biochem. Biophys. Res. Commun.* **122**, 1131 (1984).
12. C. L. Masters *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4245 (1985).
13. J. Kang *et al.*, *Nature* **325**, 733 (1987).
14. T. Iwatsubo *et al.*, *Neuron* **13**, 45 (1994).
15. T. Lührs *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17342 (2005).
16. E. Levy *et al.*, *Science* **248**, 1124 (1990).
17. A. M. Goate *et al.*, *Lancet* **i**, 352 (1989).
18. P. H. St George-Hyslop *et al.*, *Nature* **347**, 194 (1990).
19. A. M. Goate *et al.*, *Nature* **349**, 704 (1991).
20. M. C. Chartier-Harlin *et al.*, *Nature* **353**, 844 (1991).
21. J. Murrell, M. Farlow, B. Ghetti, M. D. Benson, *Science* **254**, 97 (1991).
22. C. Haass *et al.*, *Nature* **359**, 322 (1992).
23. P. Seubert *et al.*, *Nature* **359**, 325 (1992).
24. M. Shoji *et al.*, *Science* **258**, 126 (1992).
25. M. Citron *et al.*, *Nature* **360**, 672 (1992).
26. N. Suzuki *et al.*, *Science* **264**, 1336 (1994).
27. T. Wisniewski, J. Ghiso, B. Frangione, *Biochem. Biophys. Res. Commun.* **179**, 1247 (1991).
28. A. Rovelet-Lecrux *et al.*, *Nat. Genet.* **38**, 24 (2006).
29. K. Sleegers *et al.*, *Brain*, 10.1093/brain/awl203 (2006).
30. J. A. R. Nicoll *et al.*, *Neurobiol. Aging* **25**, 589 (2004).
31. D. Games *et al.*, *Nature* **373**, 523 (1995).
32. M. C. Herzig *et al.*, *Nat. Neurosci.* **7**, 954 (2004).
33. G. D. Schellenberg *et al.*, *Science* **258**, 668 (1992).
34. R. Sherrington *et al.*, *Nature* **375**, 754 (1995).
35. E. Levy-Lahad *et al.*, *Science* **269**, 973 (1995).
36. E. I. Rogaev *et al.*, *Nature* **376**, 775 (1995).
37. B. De Strooper *et al.*, *Nature* **391**, 387 (1998).
38. M. S. Wolfe *et al.*, *Nature* **398**, 513 (1999).
39. M. Citron *et al.*, *Nat. Med.* **3**, 67 (1997).
40. C. F. Lippa, L. E. Nee, H. Mori, P. St George-Hyslop, *Lancet* **352**, 1117 (1998).
41. M. J. Smith *et al.*, *Ann. Neurol.* **49**, 125 (2001).
42. B. Dermaut *et al.*, *Ann. Neurol.* **55**, 617 (2004).
43. L. E. Doglio *et al.*, *Neuron* **50**, 359 (2006).
44. R. Vassar *et al.*, *Science* **286**, 735 (1999).
45. J. Hardy, D. J. Selkoe, *Science* **297**, 353 (2002).
46. J. P. Brion, H. Passareiro, J. Nunez, J. Flament-Durand, *Arch. Biol. (Bruxelles)* **95**, 229 (1985).
47. I. Grundke-Iqbal *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4913 (1986).
48. Y. Ihara, N. Nukina, R. Miura, M. Ogawara, *J. Biochem. (Tokyo)* **99**, 1807 (1986).
49. A. Delacourte, A. Défossez, *J. Neurol. Sci.* **76**, 173 (1986).
50. K. S. Kosik, C. L. Joachim, D. J. Selkoe, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4044 (1986).
51. M. Goedert, C. M. Wischik, R. A. Crowther, J. E. Walker, A. Klug, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4051 (1988).
52. C. M. Wischik *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4506 (1988).
53. V. M.-Y. Lee, B. J. Balin, L. Otvos, J. Q. Trojanowski, *Science* **251**, 675 (1991).
54. M. Goedert, M. G. Spillantini, R. Jakes, D. Rutherford, R. A. Crowther, *Neuron* **3**, 519 (1989).
55. V. M.-Y. Lee, M. Goedert, J. Q. Trojanowski, *Annu. Rev. Neurosci.* **24**, 1121 (2001).
56. T. Revesz *et al.*, *Acta Neuropathol. (Berlin)* **97**, 170 (1999).
57. B. Ghetti *et al.*, *Neurology* **39**, 1453 (1989).
58. J. Avila, *FEBS Lett.* **580**, 2922 (2006).
59. H. Braak, E. Braak, *Acta Neuropathol. (Berlin)* **82**, 239 (1991).
60. P. Poorkaj *et al.*, *Ann. Neurol.* **43**, 815 (1998).
61. M. Hutton *et al.*, *Nature* **393**, 702 (1998).
62. M. G. Spillantini *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7737 (1998).

63. M. Hasegawa, M. J. Smith, M. Goedert, *FEBS Lett.* **437**, 207 (1998).
64. I. D'Souza *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5598 (1999).
65. G. D. J. Watts *et al.*, *Nat. Genet.* **36**, 377 (2004).
66. G. Skibinski *et al.*, *Nat. Genet.* **37**, 806 (2005).
67. M. Baker *et al.*, *Nature* **442**, 916 (2006).
68. M. Cruts *et al.*, *Nature* **442**, 920 (2006).
69. M. Baker *et al.*, *Hum. Mol. Genet.* **8**, 711 (1999).
70. H. Stefansson *et al.*, *Nat. Genet.* **37**, 129 (2005).
71. D. A. Koolen *et al.*, *Nat. Genet.* **38**, 999 (2006).
72. C. Shaw-Smith *et al.*, *Nat. Genet.* **38**, 1032 (2006).
73. A. J. Sharp *et al.*, *Nat. Genet.* **38**, 1038 (2006).
74. C. Conrad *et al.*, *Ann. Neurol.* **41**, 277 (1997).
75. E. Di Maria *et al.*, *Ann. Neurol.* **47**, 374 (2000).
76. P. Pastor *et al.*, *Ann. Neurol.* **47**, 242 (2000).
77. J. B. J. Kwok *et al.*, *Ann. Neurol.* **55**, 329 (2004).
78. J. Lewis *et al.*, *Nat. Genet.* **25**, 402 (2000).
79. B. Allen *et al.*, *J. Neurosci.* **22**, 9340 (2002).
80. J. Lewis *et al.*, *Science* **293**, 1487 (2001).
81. W. J. Strittmatter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1977 (1993).
82. D. E. Schmechel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9649 (1993).
83. R. W. Mahley, K. H. Weisgraber, Y. Huang, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5644 (2006).
84. J. L. Price, J. C. Morris, *Ann. Neurol.* **45**, 358 (1999).
85. C. G. Glabe, *Neurobiol. Aging* **27**, 570 (2006).
86. R. C. Petersen *et al.*, *Arch. Neurol.* **56**, 303 (1999).
87. R. C. Petersen *et al.*, *Arch. Neurol.* **63**, 665 (2006).
88. M. A. Mintun *et al.*, *Neurology* **67**, 446 (2006).
89. Supported by the UK Medical Research Council, the Alzheimer's Research Trust, the Parkinson's Disease Society, and the European Union Integrated Project APOPIS.

10.1126/science.1132814

100 Years and Counting: Prospects for Defeating Alzheimer's Disease

Erik D. Roberson and Lennart Mucke*

This week marks a century since the first description of Alzheimer's disease (AD). Despite approval of several drugs for AD, the disease continues to rob millions of their memories and their lives. Fortunately, many new therapies directly targeting the mechanisms underlying AD are now in the pipeline. Among the investigative AD therapies in clinical trials are several strategies to block pathogenic amyloid-β peptides and to rescue vulnerable neurons from degeneration. Complementary but less mature strategies aim to prevent the copathogenic effects of apolipoprotein E and the microtubule-associated protein tau. New insights into selective neuronal vulnerability and the link between aging and AD may provide additional entry points for therapeutic interventions. The predicted increase in AD cases over the next few decades makes the development of better treatments a matter of utmost importance and urgency.

It used to be said that neurologic diseases were easy to diagnose but impossible to treat. Today, effective treatments are available for many neurologic conditions, but for the 4.6 million new patients worldwide who will be affected by AD this year (*1*), the old mantra still rings too true. Although multiple drugs have now been approved, their expected benefits are modest. One hundred years after the discovery of AD, the lack of treatments with a major impact might be discouraging. Fortunately, basic research is identifying many of the pathways that contribute to this devastating disease (Fig. 1), providing unprecedented opportunities for the development of new treatments aimed at the root causes of AD. Here, we review several of these efforts and consider both shorter- and longer-term prospects for effectively treating AD.

Current Standard of Care

Five drugs are approved in the United States for the treatment of AD (*2*, *3*), although tacrine is now rarely used because of hepatotoxicity (Table 1). Cholinesterase inhibitors are designed to combat impairment of cholinergic neurons by slowing degradation of acetylcholine after its release at synapses. Memantine prevents overstimulation of the *N*-methyl-D-aspartate (NMDA) subtype of glutamate receptors, which may contribute to the pathogenesis of AD and other neurodegenerative conditions by causing excitotoxicity (*4*). In clinical trials, both cholinesterase inhibitors and memantine have shown beneficial but modest effects on cognitive test scores, behavioral measures, and functional outcomes (*5*–*9*). However, because the benefits of cholinesterase inhibitors are small and may be seen in only a subset of patients, their cost effectiveness has been questioned (*10*). Because memantine is beneficial in patients already taking cholinesterase inhibitors and may even reduce their side effects, the two are often used together (*9*). Many AD patients also receive antipsychotics or antidepressants to manage neuropsychiatric and behavioral symptoms or take over-the-counter preparations whose therapeutic value is uncertain, including ginkgo biloba and vitamins C and E (*2*, *11*–*14*).

In the Pipeline: Targeting Aβ

The marginal benefits of current therapies emphasize the need for more potent AD drugs. Several new compounds are now being tested for safety (phase I and IIA) and efficacy (phase IIB and III) in clinical trials (Table 2) (*15*). To date, emphasis has been on strategies to reduce the pathogenicity of amyloid-β (Aβ) peptides (*16*), widely believed to play a key role in AD.

Reducing Aβ production is one goal. Aβ is generated from the amyloid precursor protein, APP, via sequential cleavage by β- and γ-secretase (Fig. 2). γ-Secretase inhibitors have reached clinical trials, but published results are limited. One compound, LY450139, was well tolerated and reduced the amount of Aβ in the plasma, but not in the cerebrospinal fluid (CSF) (*17*). The potential for dose escalation is limited, because γ-secretase also cleaves other substrates, including Notch, and nonselective γ-secretase inhibitors have deleterious effects on embryogenesis in zebrafish and on lymphoid and gastrointestinal tissues in mammals (*18*, *19*).

Thus, several approaches are being pursued to design next-generation γ-secretase drugs that selectively reduce APP cleavage (Fig. 2). As opposed to the standard strategy of inhibiting proteases by blocking their active sites, one approach targets the substrate-docking site of γ-secretase to selectively interfere with APP binding (*20*). Another idea capitalizes on the observation that γ-secretase has an adenosine triphosphate (ATP)–binding site that selectively modulates APP processing (*21*). Blocking this site inhibits APP, but not Notch, cleavage (*22*). Yet another approach is to modulate, rather than inhibit, γ-secretase activity. Besides the γ site, γ-secretase also cleaves at a more C-terminal ε site critical for proper Notch signaling. The TMP21 accessory component of γ-secretase suppresses γ-cleavage without affecting ε-cleavage of APP or Notch, suggesting a means to inhibit Aβ production without Notch-dependent adverse effects (*23*). Lastly, even at the γ site, APP can be cleaved at different positions, creating 40– or 42–amino acid forms; the $A\beta_{42}$ peptide appears to be the most pathogenic. Certain nonsteroidal anti-inflammatory drugs (NSAIDs) allosterically modulate γ-secretase to favor production of $A\beta_{40}$ over $A\beta_{42}$ (*24*, *25*) and are now in phase III trials.

β-Secretase, whose cleavage of APP precedes that of γ-secretase (Fig. 2), is another prime target to inhibit Aβ production. It has fewer known substrates than γ-secretase and a more benign gene-knockout phenotype in mice (*26*), suggesting that β-secretase inhibitors may be safer than γ-secretase inhibitors. Genetic elimination of β-secretase prevented memory deficits in human APP transgenic mice (*27*). For structural reasons, it has been more difficult to design small-molecule inhibitors for β-secretase than for γ-secretase, but this problem appears to be surmountable (*26*). Other APP-cleaving enzymes might also be good targets. Stimulating α-secretase can reduce Aβ because the enzyme cleaves APP within Aβ (*28*–*30*). Preventing caspase cleavage of the APP intracellular domain may also be beneficial (*31*).

Gladstone Institute of Neurological Disease and Department of Neurology, University of California, San Francisco, CA 94158, USA.

*To whom correspondence should be addressed. E-mail: lmucke@gladstone.ucsf.edu

Promoting Aβ clearance by immune mechanisms is another promising approach (*32*). In a phase II trial, active immunization with $A\beta_{42}$ plus adjuvant appeared to reduce amyloid deposits in some brain regions, improve certain cognitive measures, and ameliorate CSF abnormalities in patients who developed antibodies (about 20% of those in the treatment group). However, the trial was halted because 6% of immunized patients developed meningoencephalitis (*32*). Because this complication was likely mediated by a T helper 1 (T_H1)–cell response, much effort has been made to circumvent that arm of the immune response while preserving the beneficial effect of antibodies against Aβ (anti-Aβ) on amyloid removal by microglia and blood-derived macrophages (*32*).

One approach is passive immunization with anti-Aβ, avoiding the T cell response (*33*). Although passive immunization resulted in cerebral microhemorrhages in some human APP transgenic mouse lines (*34*), antibody deglycosylation circumvents this complication (*35*). An ongoing phase IIA trial of passive immunization has progressed to an advanced stage without interruptions due to adverse events.

Active immunization with Aβ may still be useful too. In APP transgenic mice, immunization with the $A\beta_{1-15}$ fragment (*36*) or with Aβ coupled to carrier protein (*32*) led to amyloid clearance without activating the undesired T cell response. Delivering Aβ through the nasal mucosa may also avoid adverse T cell effects seen with intramuscular injections (*37*). However, given the complications of the original immunization trial, the utmost caution is required in extrapolating these results to humans. Interestingly, the immune-modulatory polypeptide glatiramer effectively cleared amyloid deposits in APP mice independently of anti-Aβ (*38*). This drug is already used to treat multiple sclerosis and will likely soon be tested in AD patients.

Lastly, certain small molecules disrupt Aβ aggregation. Although published data are limited, such compounds show promise in animal models (*39*, *40*) and are in phase III clinical trials.

Clinical Trials Beyond Aβ

Not all current clinical trials are aimed at Aβ. Because depletion of nerve growth factor (NGF) may contribute to loss of cholinergic neurons in AD, boosting NGF has been pursued using several strategies (*41*). In the boldest procedure, fibroblasts are isolated from AD patients, transduced with an NGF-encoding viral vector, and implanted stereotactically into the forebrains of patients (*42*). Although only eight subjects were tested in the phase I trial, the procedure was reasonably well tolerated when carried out under full anesthesia and may have improved cognitive performance and cerebral perfusion.

Several "off-the-shelf" drugs are also being tested for efficacy in AD. For example, NSAIDs (*43*) and cholesterol-lowering statins (*44*) were associated with decreased risk of developing AD in retrospective series. Prospective trials, though, have been mostly disappointing. Some interpret the negative prospective trials as too little too late and argue for retesting in prevention trials. Others suspect that the negative trials reflect the weaknesses of retrospective data, e.g., that other factors distinguish NSAID users from nonusers and may be responsible for the differential AD risk. Because prevention trials are generally even larger and more expensive than treatment trials, this issue has major implications for resource allocation in the field, and debate remains about whether funds would be better devoted to supporting more basic research and developing more effective drugs.

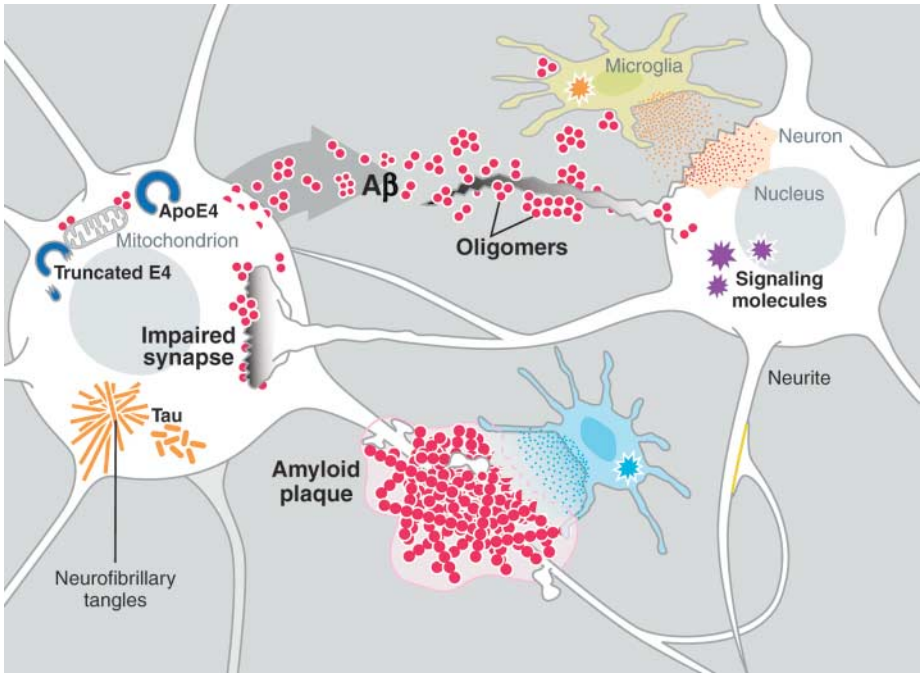


Fig. 1. Molecular and cellular processes presumed to participate in AD pathogenesis. Aβ peptides produced by neurons and other brain cells aggregate into a variety of assemblies, some of which impair synapses and neuronal dendrites, either directly or through the engagement of glial loops. Build-up of pathogenic Aβ assemblies could result from increased production or aggregation or from deficient clearance mechanisms. ApoE4 and tau promote Aβ-induced neuronal injury and also have independent adverse effects. Microglia could be beneficial or harmful, depending on which of their signaling cascades and functions are engaged. This multifactorial scenario leads to progressive disintegration of neural circuits, isolation and loss of neurons, network failure, and neurological decline.

Table 1. Food and Drug Administration–approved treatments for AD.

Drug	Approved for
Cholinesterase inhibitors	
Donepezil	Mild to moderate AD
Galantamine	Mild to moderate AD
Rivastigmine	Mild to moderate AD
Tacrine	Mild to moderate AD
NMDA receptor antagonist	
Memantine	Moderate to severe AD

Table 2. Selected treatments in clinical trials for AD. For more information on these and other trials, see (*15*, *89*, *90*).

Treatment strategies
Phase III
Aβ aggregation inhibitors
Antioxidants
γ-Secretase modulators
NGF mimics
PPARγ agonists
HMG-CoA reductase inhibitors (statins)
Phase II
Ampakines
Calcium channel blockers
GABA receptor antagonists
γ-Secretase inhibitors
Glycogen synthase kinase inhibitors
Intravenous immunoglobulin
Muscarinic receptor agonists
New cholinesterase inhibitors
Nicotinic receptor modulators
Passive Aβ immunization
Phosphodiesterase inhibitors
Serotonin receptor antagonists
Phase I
Active Aβ immunization
NGF gene therapy

The pleiotropic actions of NSAIDs and statins complicate matters further. For example, some NSAIDs have anti-Aβ as well as anti-inflammatory properties (*43*). Notably, most negative prospective NSAID trials tested only selective cyclooxygenase 2 (COX-2) inhibitors (to reduce gastrointestinal side effects), but these drugs lack the anti-Aβ effects of other NSAIDs and may even increase Aβ production (*43*, *45*). Whether optimized NSAIDs that safely combine anti-Aβ and anti-inflammatory activities will be more efficacious remains to be determined.

Lastly, several drugs have been developed to ameliorate AD-related abnormalities by modulating various neurotransmitter receptors (Table 2) (*15*). Some provide primarily symptomatic benefits and others may directly target AD pathogenesis, but their effectiveness in AD remains unclear (*15*).

Neglected Opportunities

The apolipoprotein E (*APOE*) ε4 allele has emerged as the major genetic risk factor for AD, whereas individuals with ε2/ε2 or ε3/ε2 genotypes rarely develop the disease (*46*, *47*). Nature seems to be suggesting an important avenue toward treating this disease, but specific strategies had been lacking. However, promising new approaches to counteracting the adverse effects of apoE4 or leveraging the beneficial effects of apoE2 or apoE3 are beginning to emerge from basic research (*48*, *49*).

The two domains of apoE interact more closely with each other in apoE4 than in the other apoE isoforms, which may account for many of apoE4's adverse effects (*48*, *49*). Compounds that disrupt domain interaction, inducing apoE4 to adopt a more beneficial structure and function, are being developed. In addition, selective pathogenic cleavage of apoE4 yields a truncated apoE that may impair mitochondrial energy production and disrupt the cytoskeleton (Fig. 1). When identified, the putative apoE-cleaving enzyme may be an attractive drug target (*49*). Complementary efforts exploit the differential effects of apoE isoforms on the formation and clearance of amyloid deposits (*50*–*52*).

Another molecule that may have been inadvertently overshadowed by Aβ is the microtubule-associated protein tau, the main constituent of neurofibrillary tangles. Tau undergoes many AD-related posttranslational modifications (*53*). Tau phosphorylation increases dramatically in AD (*54*), suggesting tau kinase inhibitors as an AD treatment. Lithium, which inhibits tau phosphorylation with beneficial effects in animal models, is in clinical trials. However, potential redundancy between the many kinases that phosphorylate tau and uncertainty about which phosphorylation events are truly pathogenic raise challenging issues in the design of tau kinase inhibitors. Other approaches to tau include blocking its aggregation, either directly (*55*) or by inhibiting its proteolysis (*56*). Because hyperphosphorylated tau tends to dissociate from microtubules, reducing their stability, microtubule-stabilizing drugs represent an alternative approach (*57*).

Another approach to apoE, tau, and even APP may be reducing their overall amounts without targeting particular posttranslational modifications; because mild (~20%) overexpression of APP or tau increases AD risk (*58*, *59*), reducing their overall concentrations by a similar magnitude might decrease AD risk. Even complete ablation of these proteins in knockout mice is rather well tolerated (*60*–*62*). Small molecules that suppress APP or tau concentrations have already been identified (*63*, *64*). Proof-of-principle studies are needed to determine whether such strategies will prevent cognitive impairment in animal models of AD.

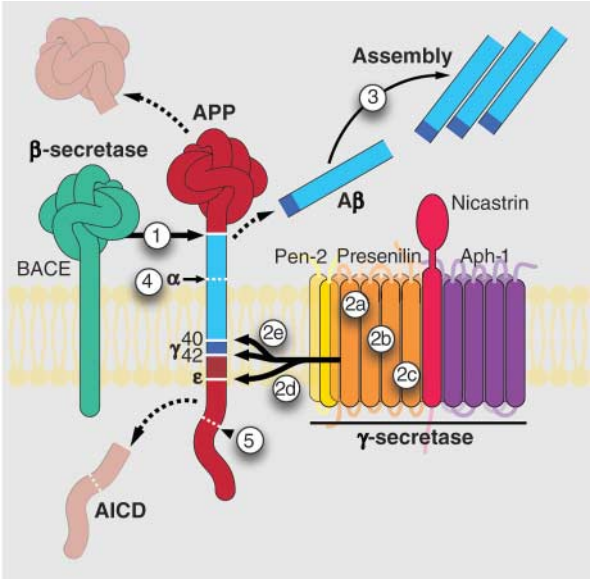


Fig. 2. Drug targets involved in Aβ production and assembly. Aβ production depends on sequential proteolytic cleavage of APP by β-secretase (marked 1), also known as β-site APP-cleaving enzyme 1 (BACE1), and the multiprotein γ-secretase complex (2) (*88*). γ-Secretase targets include the enzyme's active site, substrate docking site, and ATP-binding site (2a to 2c) and its predilection for γ- versus ε-cleavage (2d) or for generating $Aβ_{42}$ versus $Aβ_{40}$ (2e). The pathogenicity of released Aβ peptides depends on self-assembly (3). APP cleavage by α-secretases (ADAM family metalloproteases) prevents Aβ production (4). Caspase cleavage of the APP intracellular domain (AICD) generates a C31 fragment (5) that may participate in Aβ-induced toxicity or act independently.

Unresolved Questions with Therapeutic Implications

Basic research has paved the way to new drugs in clinical trials, but several fundamental questions about AD remain. Resolving these issues could usher in the next generation of therapeutics.

The selective vulnerability of specific neuronal populations to AD is one such issue. It is unclear whether this vulnerability relates to cell-autonomous structural, biochemical, or electrophysiological properties; to interactions with glial cells; or to the extracellular milieu in specific brain regions. Recent studies have added yet another possibility to the list: differences in neuronal activities within distinct neural networks.

There is striking overlap between the anatomic regions most affected by AD and the "default-mode" network, active when the mind is focused on nothing in particular (*65*, *66*). Because neuronal activity increases the production and release of Aβ, excessive activity in the default-mode network might make it vulnerable to AD (*66*–*68*). This pathogenic interaction may also explain, at least in part, the beneficial effects of mental activity and environmental enrichment (*69*, *70*), which should decrease activity and, thus, Aβ in this network. Rapid fluctuations of cognitive function that cannot be explained by neuronal loss raise additional possibilities for the therapeutic modulation of network activities (*71*).

Another important question relates to the link between aging and neurodegenerative disorders. Can it be manipulated to prevent or delay AD? A cadre of peptidases is essential to clear Aβ from the brain (*72*–*74*), and their activity appears to decline with age, possibly contributing to AD (*74*–*76*). Boosting Aβ-degrading enzymes, such as neprilysin (*77*, *78*), endothelin-converting enzyme (*79*), or cathepsin B (*74*), protects transgenic mice from Aβ, highlighting their therapeutic potential. Because it is difficult to pharmacologically activate these enzymes, the best way to leverage their Aβ-degrading activities may be through druggable factors that regulate them naturally (*79*, *80*).

Oxidative stress is a widely explored link between aging and neurodegenerative diseases (*4*, *14*). Given the ample evidence for oxidative stress in AD, it is surprising that trials of antioxidants, such as vitamins C and E, have yielded mostly disappointing results (*12*, *14*). Have the right compounds not yet been tested? Trials of other antioxidants such as coenzyme Q10 and curcumin are ongoing. Or might it be better to focus on activating intrinsic defense systems, for example, through caloric restriction, which reduces oxidative stress, delays aging, and may lower AD risk (*81*, *82*)?

Stem cells are often touted as a potential AD treatment. However, even beyond the much-discussed ethical and political hurdles, there are important scientific questions about potential use of stem cells for AD (*83*–*85*). For example, AD affects different types of neurons in multiple brain regions; how many must be replaced, and

can stem cells differentiate into all the necessary populations? Would stem cell grafts integrate both structurally and functionally into vulnerable neuronal networks? Or might these grafts provide benefits through integration-independent effects, such as neurotrophin release? In either case, will the aged brain support their therapeutic activities? Moreover, will their treatment capacity be limited in the milieu of Aβ, apoE4, tau, and inflammatory mediators found in the AD brain? Definitive answers to these questions are needed but may not become available for many years.

Lastly, when exactly does AD begin, and how early will one have to intervene with pathogenic mechanisms to prevent its clinical manifestations? More and more emphasis is being placed on early detection, based on the reasonable assumption that AD will be easier to prevent than reverse. Whether earlier really is better will depend, in large part, on the safety and side effect profiles of emerging AD treatments, as highlighted by ongoing debates about the best timing for treatments of other chronic conditions, such as HIV infection and Parkinson's disease (*86*, *87*).

Conclusions

When the secretases that produce Aβ were first identified in the late 1990s, some people felt that the writing was on the wall for AD: A little effort on protease inhibitor development and the end was near. Today, most are substantially more circumspect, and there seems to be consensus that multiple drugs will be required. For one, the field now recognizes the important pathogenic roles of molecules beyond Aβ. Indeed, it seems likely that, for example, an AD patient with a genetic mutation causing Aβ overproduction and an AD patient with two *APOE* ε4 alleles plus a history of head trauma might benefit from different regimens, one focused on lowering Aβ and the other including apoE-targeting drugs and neuroprotectants.

The need for drugs with different modes of action and for individualized regimens creates imposing challenges. Clinical trials must carefully consider their inclusion and exclusion criteria, and subgroup analysis becomes critical. Better synergy between industry and academia is required to speed the transition from target identification to drug development. Funding must increase to reduce the mismatch between the rapidly growing economic threat from AD and the limited resources available to fight it.

Notwithstanding these challenges, there is good reason for optimism. With the many exciting prospects now in the pipeline and the steady flow of insights into disease pathogenesis from basic laboratories, the arsenal of clinicians fighting AD should be more fully stocked at the next major anniversary of Alzheimer's discovery.

References and Notes

1. C. P. Ferri *et al.*, *Lancet* **366**, 2112 (2005).
2. P. M. Doraiswamy, G. L. Xiong, *Expert Opin. Pharmacother.* **7**, 1 (2006).
3. A. Lleó, S. M. Greenberg, J. H. Growdon, *Annu. Rev. Med.* **57**, 513 (2006).
4. M. P. Mattson, *Nature* **430**, 631 (2004).
5. C. W. Ritchie, D. Ames, T. Clayton, R. Lai, *Am. J. Geriatr. Psychiatry* **12**, 358 (2004).
6. A. Whitehead *et al.*, *Int. J. Geriatr. Psychiatry* **19**, 624 (2004).
7. J. Birks, *Cochrane Database Syst. Rev.* 10.1002/14651858.CD005593 (2006).
8. R. Bullock, *Alzheimer Dis. Assoc. Disord.* **20**, 23 (2006).
9. R. McShane, A. Areosa Sastre, N. Minakaran, *Cochrane Database Syst. Rev.* 10.1002/14651858.CD003154 (2006).
10. M. Maggini, N. Vanacore, R. Raschetti, *PLoS Med.* **3**, e140 (2006).
11. M. R. Farlow, *Expert Rev. Neurotherapeutics* **4**, 799 (2004).
12. L. A. Boothby, P. L. Doering, *Ann. Pharmacother.* **39**, 2073 (2005).
13. E. R. Miller III *et al.*, *Ann. Intern. Med.* **142**, 37 (2005).
14. A. Nunomura *et al.*, *J. Neuropathol. Exp. Neurol.* **65**, 631 (2006).
15. J. S. Jacobsen, P. Reinhart, M. N. Pangalos, *NeuroRx* **2**, 612 (2005).
16. M. Citron, *Nat. Rev. Neurosci.* **5**, 677 (2004).
17. E. R. Siemers *et al.*, *Neurology* **66**, 602 (2006).
18. A. Geling, H. Steiner, M. Willem, L. Bally-Cuif, C. Haass, *EMBO Rep.* **3**, 688 (2002).
19. G. T. Wong *et al.*, *J. Biol. Chem.* **279**, 12876 (2004).
20. M. S. Wolfe, *Biochemistry* **45**, 7931 (2006).
21. P. C. Fraering *et al.*, *J. Biol. Chem.* **280**, 41987 (2005).
22. W. J. Netzer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12444 (2003).
23. F. Chen *et al.*, *Nature* **440**, 1208 (2006).
24. J. L. Eriksen *et al.*, *J. Clin. Invest.* **112**, 440 (2003).
25. A. Lleó *et al.*, *Nat. Med.* **10**, 1065 (2004).
26. M. Citron, *Trends Pharmacol. Sci.* **25**, 92 (2004).
27. M. Ohno *et al.*, *Eur. J. Neurosci.* **23**, 251 (2006).
28. R. Etcheberrigaray *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11141 (2004).
29. A. Caccamo *et al.*, *Neuron* **49**, 671 (2006).
30. W. Qin *et al.*, *J. Biol. Chem.* **281**, 21745 (2006).
31. V. Galvan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7130 (2006).
32. H. L. Weiner, D. Frenkel, *Nat. Rev. Immunol.* **6**, 404 (2006).
33. F. Bard *et al.*, *Nat. Med.* **6**, 916 (2000).
34. M. Pfeifer *et al.*, *Science* **298**, 1379 (2002).
35. D. M. Wilcock *et al.*, *J. Neurosci.* **26**, 5340 (2006).
36. M. Maier *et al.*, *J. Neurosci.* **26**, 4717 (2006).
37. H. L. Weiner *et al.*, *Ann. Neurol.* **48**, 567 (2000).
38. D. Frenkel, R. Maron, D. S. Burt, H. L. Weiner, *J. Clin. Invest.* **115**, 2423 (2005).
39. F. Gervais *et al.*, *Neurobiol. Aging* 10.1016/j.neurobiolaging.2006.02.015 (2006).
40. J. McLaurin *et al.*, *Nat. Med.* **12**, 801 (2006).
41. M. H. Tuszynski, *Lancet Neurol.* **1**, 51 (2002).
42. M. H. Tuszynski *et al.*, *Nat. Med.* **11**, 551 (2005).
43. K. P. Townsend, D. Praticò, *FASEB J.* **19**, 1592 (2005).
44. B. Wolozin, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5371 (2001).
45. T. Kukar *et al.*, *Nat. Med.* **11**, 545 (2005).
46. L. A. Farrer *et al.*, *JAMA* **278**, 1349 (1997).
47. J. Raber, Y. Huang, J. W. Ashford, *Neurobiol. Aging* **25**, 641 (2004).
48. D. M. Hatters, C. A. Peters-Libeu, K. H. Weisgraber, *Trends Biochem. Sci.* **31**, 445 (2006).
49. R. W. Mahley, K. H. Weisgraber, Y. Huang, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5644 (2006).
50. M. Koistinaho *et al.*, *Nat. Med.* **10**, 719 (2004).
51. M. Sadowski *et al.*, *Am. J. Pathol.* **165**, 937 (2004).
52. J. C. Dodart *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1211 (2005).
53. C. X. Gong, F. Liu, I. Grundke-Iqbal, K. Iqbal, *J. Neural Transm.* **112**, 813 (2005).
54. W. H. Stoothoff, G. V. Johnson, *Biochim. Biophys. Acta* **1739**, 280 (2005).
55. I. Khlistunova *et al.*, *J. Biol. Chem.* **281**, 1205 (2006).
56. C. W. Cotman, W. W. Poon, R. A. Rissman, M. Blurton-Jones, *J. Neuropathol. Exp. Neurol.* **64**, 104 (2005).
57. B. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 227 (2005).
58. A. J. Myers *et al.*, *Hum. Mol. Genet.* **14**, 2399 (2005).
59. N. Brouwers *et al.*, *Brain* 10.1093/brain/awl212 (2006).
60. S. H. Zhang, R. L. Reddick, J. A. Piedrahita, N. Maeda, *Science* **258**, 468 (1992).
61. A. Harada *et al.*, *Nature* **369**, 488 (1994).
62. H. Zheng *et al.*, *Cell* **81**, 525 (1995).
63. S. Tucker *et al.*, *Curr. Alzheimer Res.* **3**, 221 (2006).
64. C. A. Dickey *et al.*, *FASEB J.* **20**, 753 (2006).
65. M. D. Greicius, G. Srivastava, A. L. Reiss, V. Menon, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4637 (2004).
66. R. L. Buckner *et al.*, *J. Neurosci.* **25**, 7709 (2005).
67. F. Kamenetz *et al.*, *Neuron* **37**, 925 (2003).
68. J. R. Cirrito *et al.*, *Neuron* **48**, 913 (2005).
69. L. Fratiglioni, S. Paillard-Borg, B. Winblad, *Lancet Neurol.* **3**, 343 (2004).
70. O. Lazarov *et al.*, *Cell* **120**, 701 (2005).
71. J. J. Palop, J. Chin, L. Mucke, *Nature* **443**, 768 (2006).
72. N. Iwata *et al.*, *Nat. Med.* **6**, 143 (2000).
73. D. J. Selkoe, *Neuron* **32**, 177 (2001).
74. S. Mueller-Steiner *et al.*, *Neuron* **51**, 703 (2006).
75. N. Iwata, Y. Takaki, S. Fukami, S. Tsubuki, T. C. Saido, *J. Neurosci. Res.* **70**, 493 (2002).
76. A. Caccamo, S. Oddo, M. C. Sugarman, Y. Akbari, F. M. Laferla, *Neurobiol. Aging* **26**, 645 (2005).
77. M. A. Leissring *et al.*, *Neuron* **40**, 1087 (2003).
78. R. A. Marr *et al.*, *J. Neurosci.* **23**, 1992 (2003).
79. D.-S. Choi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8215 (2006).
80. T. Saito *et al.*, *Nat. Med.* **11**, 434 (2005).
81. R. S. Sohal, R. Weindruch, *Science* **273**, 59 (1996).
82. J. A. Luchsinger, M. X. Tang, S. Shea, R. Mayeux, *Arch. Neurol.* **59**, 1258 (2002).
83. G. M. Bernal, D. A. Peterson, *Aging Cell* **3**, 345 (2004).
84. G. Martino, S. Pluchino, *Nat. Rev. Neurosci.* **7**, 395 (2006).
85. O. Lindvall, Z. Kokaia, *Nature* **441**, 1094 (2006).
86. S. G. Deeks, *Br. Med. J.* **332**, 1489 (2006).
87. A. H. V. Schapira, J. Obeso, *Ann. Neurol.* **59**, 559 (2006).
88. D. Edbauer *et al.*, *Nat. Cell Biol.* **5**, 486 (2003).
89. National Library of Medicine, National Institutes of Health, www.clinicaltrials.gov.
90. Alzheimer Research Forum, www.alzforum.org/dis/tre/drc/.

91. We thank L. Gan and Y. Huang for helpful discussions, J. Carroll for graphics, G. Howard and S. Ordway for editorial review, and D. McPherson and L. Manuntag for administrative assistance. Supported by a Stephen D. Bechtel Jr. Young Investigator Award (E.D.R.) and NIH grants NS054811 (E.D.R), AG011385, and AG022074 (L.M.). L.M. has received consulting fees from Merck and honoraria or speaking fees from Amgen, Elan, and Pfizer. E.R. has received consulting fees from Rinat Neuroscience.

10.1126/science.1132813

Accelerated Evolution of Conserved Noncoding Sequences in Humans

Shyam Prabhakar,[1,2]* James P. Noonan,[1,2]* Svante Pääbo,[3] Edward M. Rubin[1,2]†

Traits that distinguish humans from other primates originated in human-specific DNA sequence changes. To investigate whether gene regulatory or other functional noncoding elements in the human genome bear the signature of accelerated evolution, we determined the occurrence of human-specific substitutions in 110,549 conserved noncoding sequences (CNSs) previously identified by whole-genome sequence comparisons (*1*).

To identify CNSs that accelerated in the human lineage, we developed a test statistic that evaluated the likelihood of observing the configuration of human-specific substitutions present in a given CNS. We assigned each CNS a human-acceleration *P* value based on the probability of observing a configuration of equal or smaller likelihood under the null model of constrained evolution (*1*). We identified 992 elements with a significant excess of human-specific substitutions at $P \leq 0.005$, 79% more than we would expect to see by chance (Fig. 1A and table S1).

To ascertain whether accelerated CNSs disproportionately occur near genes with particular functions, we obtained the Gene Ontology (GO) annotations of the closest neighboring RefSeq gene for all 110,549 CNSs and assigned those annotations to each CNS. Our method explicitly accounted for the unequal distribution of CNSs among GO categories. We then sought to identify GO terms with a significant excess of accelerated CNSs after correcting for variation in statistical power across CNSs (*1*). *P* value thresholds were set to adjust for multiple testing.

The cellular component GO term most significantly enriched in accelerated CNSs was basal lamina (Fig. 1B and table S1). Of the 12 accelerated CNSs in this category, 9 were associated with the dystrophin-associated glycoprotein complex, disruptions of which cause muscle and neuronal diseases (*2*, *3*). Cell adhesion was the only biological process displaying a significant excess of CNSs accelerated in human (Fig. 1B and table S1). Many of these CNSs were associated with genes involved in neuronal cell adhesion, such as cadherins and protocadherins, contactins, and neuroligins. To quantify this observation, we constructed a composite neuronal adhesion GO term by intersecting GO cell adhesion genes with genes annotated in the Entrez Gene database as having evidence of neuronal function. We chose this approach because neuronal adhesion genes are poorly annotated in GO. We found a significant excess of accelerated CNSs neighboring genes with both GO cell adhesion and Entrez Gene neuronal annotations [$P = 0.0069$, one-tailed power-adjusted exact test (*1*) (Fig. 1C and table S1)]. However, when these overlapping accelerated CNSs were removed from the analysis, the number of accelerated CNSs with only GO cell adhesion or Entrez Gene neuronal function annotations was not significantly greater than expected. Thus, the strongest signal of human-specific noncoding sequence evolution that we detected was near genes specifically involved in neuronal cell adhesion.

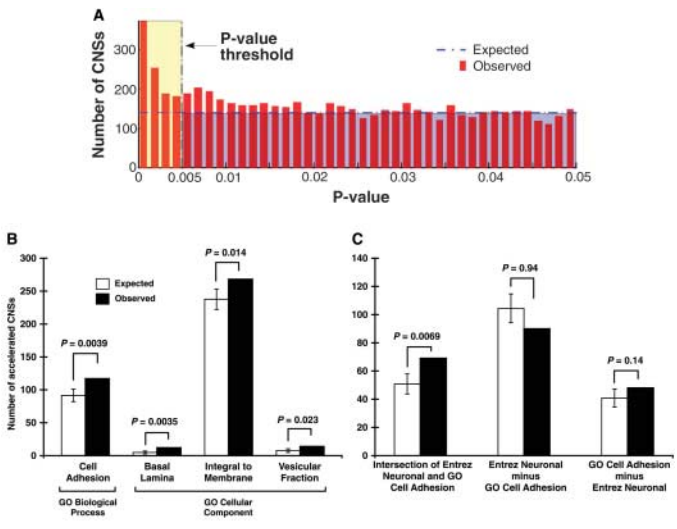


Fig. 1. (**A**) Numbers of human-specific substitution observed versus the uniform distribution expected by random chance. (**B**) GO biological process and cellular component terms enriched in accelerated CNSs. Enrichment *P* values were calculated with the power-adjusted exact test (one-tailed) (*1*). (**C**) Human-accelerated CNSs are disproportionately associated with genes functioning specifically in neuronal cell adhesion. Error bars indicate 1σ.

To determine whether the pattern of CNS acceleration in humans was recapitulated in other lineages, we identified accelerated CNSs in chimpanzee and mouse (*1*). We observed 1050 accelerated CNSs in chimpanzee, only 34 of which were also accelerated in human, indicating a general lack of overlap (table S1). However, CNSs accelerated in chimpanzee were also significantly enriched near neuronal cell adhesion genes (expected = 54, observed = 70, $P = 0.018$). This suggests independent accelerated evolution of neuronal cell adhesion functions in the human and chimpanzee lineages. For the 4607 CNSs accelerated in mouse, we did not detect any enrichment near neuronal cell adhesion genes ($P = 0.99$).

It is unlikely that acceleration of neuronal adhesion CNSs in humans and chimpanzees resulted in the same neuronal phenotypes, because the CNSs accelerated in the two lineages are largely disjoint and would therefore have had different consequences for brain development and cognitive function. Our findings do suggest that cis-regulatory and other noncoding changes may have contributed to the modifications in brain development and function that gave rise to uniquely human cognitive traits.

References and Notes

1. Materials and methods are available as supporting material on *Science* Online.
2. I. Dalkilic, L. M. Kunkel, *Curr. Opin. Genet. Dev.* **13**, 231 (2003).
3. M. P. Moizard *et al.*, *Eur. J. Hum. Genet.* **8**, 552 (2000).
4. Supported by grant HL066681 to the Berkeley PGA, NIH Programs for Genomic Application, funded by the National Heart, Lung, and Blood Institute, and by the director, Office of Science, Office of Basic Energy Sciences, of the U.S. DOE under Contract No. DE-AC02-05CH11231. J.P.N. was supported by NIH National Research Service Award fellowship 1-F32-GM074367.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/786/DC1
Materials and Methods
Table S1

1 June 2006; accepted 14 August 2006
10.1126/science.1130738

[1]U.S. Department of Energy (DOE) Joint Genome Institute, Walnut Creek, CA 94598, USA. [2]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: emrubin@lbl.gov

Impacts of Biodiversity Loss on Ocean Ecosystem Services

Boris Worm,[1]* Edward B. Barbier,[2] Nicola Beaumont,[3] J. Emmett Duffy,[4] Carl Folke,[5,6] Benjamin S. Halpern,[7] Jeremy B. C. Jackson,[8,9] Heike K. Lotze,[1] Fiorenza Micheli,[10] Stephen R. Palumbi,[10] Enric Sala,[8] Kimberley A. Selkoe,[7] John J. Stachowicz,[11] Reg Watson[12]

Human-dominated marine ecosystems are experiencing accelerating loss of populations and species, with largely unknown consequences. We analyzed local experiments, long-term regional time series, and global fisheries data to test how biodiversity loss affects marine ecosystem services across temporal and spatial scales. Overall, rates of resource collapse increased and recovery potential, stability, and water quality decreased exponentially with declining diversity. Restoration of biodiversity, in contrast, increased productivity fourfold and decreased variability by 21%, on average. We conclude that marine biodiversity loss is increasingly impairing the ocean's capacity to provide food, maintain water quality, and recover from perturbations. Yet available data suggest that at this point, these trends are still reversible.

What is the role of biodiversity in maintaining the ecosystem services on which a growing human population depends? Recent surveys of the terrestrial literature suggest that local species richness may enhance ecosystem productivity and stability (*1*–*3*). However, the importance of biodiversity changes at the landscape level is less clear, and the lessons from local experiments and theory do not seem to easily extend to long-term, large-scale management decisions (*3*). These issues are particularly enigmatic for the world's oceans, which are geographically large and taxonomically complex, making the scaling up from local to global scales potentially more difficult (*4*). Marine ecosystems provide a wide variety of goods and services, including vital food resources for millions of people (*5*, *6*). A large and increasing proportion of our population lives close to the coast; thus the loss of services such as flood control and waste detoxification can have disastrous consequences (*7*, *8*). Changes in marine biodiversity are directly caused by exploitation, pollution, and habitat destruction, or indirectly through climate change and related perturbations of ocean biogeochemistry (*9*–*13*). Although marine extinctions are only slowly uncovered at the global scale (*9*), regional ecosystems such as estuaries (*10*), coral reefs (*11*), and coastal (*12*) and oceanic fish communities (*13*) are rapidly losing populations, species, or entire functional groups. Although it is clear that particular species provide critical services to society (*6*), the role of biodiversity per se remains untested at the ecosystem level (*14*). We analyzed the effects of changes in marine biodiversity on fundamental ecosystem services by combining available data from sources ranging from small-scale experiments to global fisheries.

Experiments. We first used meta-analysis of published data to examine the effects of variation in marine diversity (genetic or species richness) on primary and secondary productivity, resource use, nutrient cycling, and ecosystem stability in 32 controlled experiments. Such effects have been contentiously debated, particularly in the marine realm, where high diversity and connectivity may blur any deterministic effect of local biodiversity on ecosystem functioning (*1*). Yet when the available experimental data are combined (*15*), they reveal a strikingly general picture (Fig. 1). Increased diversity of both primary producers (Fig. 1A) and consumers (Fig. 1B) enhanced all examined ecosystem processes. Observed effect sizes corresponded to a 78 to 80% enhancement of primary and secondary production in diverse mixtures relative to monocultures and a 20 to 36% enhancement of resource use efficiency (Fig. 1, A and B).

Experiments that manipulated species diversity (Fig. 1B) or genetic diversity (Fig. 1C) both found that diversity enhanced ecosystem stability, here defined as the ability to withstand recurrent perturbations. This effect was linked

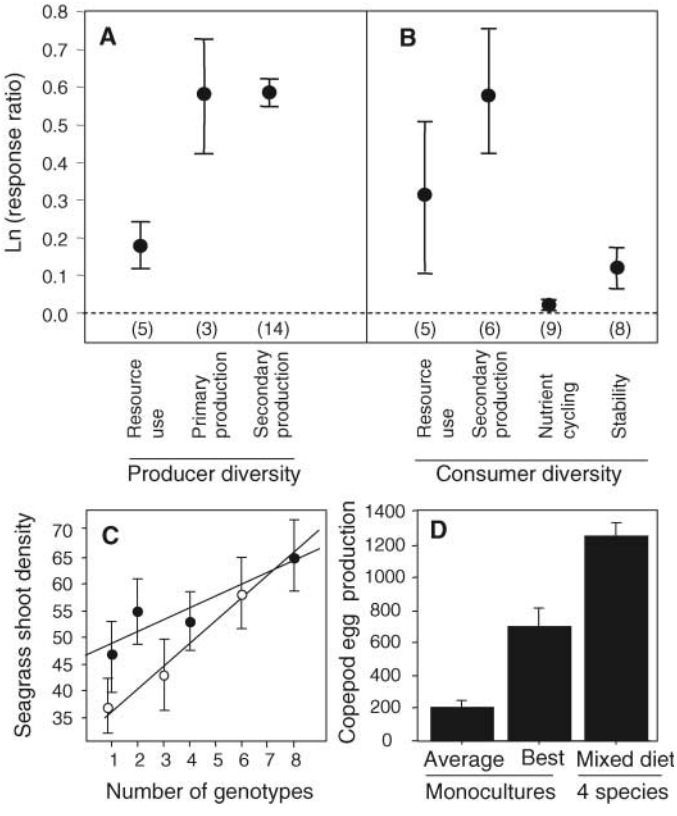


Fig. 1. Marine biodiversity and ecosystem functioning in controlled experiments. Shown are response ratios [ln(high/low diversity) ±95% confidence interval (CI)] of ecosystem processes to experimental manipulations of species diversity of (**A**) primary producers (plants and algae), and (**B**) consumers (herbivores and predators). Increased diversity significantly enhanced all examined ecosystem functions ($0.05 > P > 0.0001$). The number of studies is given in parentheses. (**C**) Genetic diversity increased the recovery of seagrass ecosystems after overgrazing (solid circles) and climatic extremes (open circles). (**D**) Diet diversity enhanced reproductive capacity in zooplankton over both the average- and best-performing monocultures.

[1]Department of Biology, Dalhousie University, Halifax, NS, Canada B3H 4J1. [2]Department of Economics and Finance, University of Wyoming, Laramie, WY 82071, USA. [3]Plymouth Marine Laboratory, Plymouth PL1 3DH, UK. [4]Virginia Institute of Marine Sciences, Gloucester Point, VA 23062–1346, USA. [5]Department of Systems Ecology, Stockholm University, Stockholm, SE-106 91 Sweden. [6]Beijer International Institute of Ecological Economics, Royal Swedish Academy of Sciences, SE-104 05, Stockholm, Sweden. [7]National Center for Ecological Analysis and Synthesis, Santa Barbara, CA 93101, USA. [8]Center for Marine Biodiversity and Conservation, Scripps Institution of Oceanography, La Jolla, CA 92093–0202, USA. [9]Smithsonian Tropical Research Institute, Box 2072, Balboa, Republic of Panama. [10]Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950, USA. [11]Section of Evolution and Ecology, University of California, Davis, CA 95616, USA. [12]Fisheries Centre, University of British Columbia, Vancouver, BC, Canada V6T 1Z4.

*To whom correspondence should be addressed. E-mail: bworm@dal.ca

to either increased resistance to disturbance (*16*) or enhanced recovery afterward (*17*). A number of experiments on diet mixing further demonstrated the importance of diverse food sources for secondary production and the channeling of that energy to higher levels in the food web (Fig. 1D). Different diet items were required to optimize different life-history processes (growth, survival, and fecundity), leading to maximum total production in the mixed diet. In summary, experimental results indicate robust positive linkages between biodiversity, productivity, and stability across trophic levels in marine ecosystems. Identified mechanisms from the original studies include complementary resource use, positive interactions, and increased selection of highly performing species at high diversity.

Coastal ecosystems. To test whether experimental results scale up in both space and time, we compiled long-term trends in regional biodiversity and services from a detailed database of 12 coastal and estuarine ecosystems (*10*) and other sources (*15*). We examined trends in 30 to 80 (average, 48) economically and ecologically important species per ecosystem. Records over the past millennium revealed a rapid decline of native species diversity since the onset of industrialization (Fig. 2A). As predicted by experiments, systems with higher regional species richness appeared more stable, showing lower rates of collapse and extinction of commercially important fish and invertebrate taxa over time (Fig. 2B, linear regression, $P < 0.01$). Overall, historical trends led to the present depletion (here defined as >50% decline over baseline abundance), collapse (>90% decline), or extinction (100% decline) of 91, 38, or 7% of species, on average (Fig. 2C). Only 14% recovered from collapse (Fig. 2C); these species were mostly protected birds and mammals.

These regional biodiversity losses impaired at least three critical ecosystem services (Fig. 2D): number of viable (noncollapsed) fisheries (–33%); provision of nursery habitats such as oyster reefs, seagrass beds, and wetlands (–69%); and filtering and detoxification services provided by suspension feeders, submerged vegetation, and wetlands (–63%). Loss of filtering services probably contributed to declining water quality (*18*) and the increasing occurrence of harmful algal blooms, fish kills, shellfish and beach closures, and oxygen depletion (Fig. 2E). Increasing coastal flooding events (Fig. 2E) are linked to sea level rise but were probably accelerated by historical losses of floodplains and erosion control provided by coastal wetlands, reefs, and submerged vegetation (*7*). An increased number of species invasions over time (Fig. 2E) also coincided with the loss of native biodiversity; again, this is consistent with experimental results (*19*). Invasions did not compensate for the loss of native biodiversity and services, because they comprised other species groups, mostly microbial, plankton, and small invertebrate taxa (*10*). Although causal relationships are difficult to infer, these data suggest that substantial loss of biodiversity (Fig. 2, A and C) is closely associated with regional loss of ecosystem services (Fig. 2D) and increasing risks for coastal inhabitants (Fig. 2E). Experimentally derived predictions that more species-rich systems should be more stable in delivering services (Fig. 1) are also supported at the regional scale (Fig. 2B).

Large marine ecosystems. At the largest scales, we analyzed relationships between biodiversity and ecosystem services using the global catch database from the United Nations Food and Agriculture Organization (FAO) and other sources (*15*, *20*). We extracted all data on fish and invertebrate catches from 1950 to 2003 within all 64 large marine ecosystems (LMEs) worldwide. LMEs are large (>150,000 km^2) ocean regions reaching from estuaries and coastal areas to the seaward boundaries of continental shelves and the outer margins of the major current systems (*21*). They are characterized by distinct bathymetry, hydrography, productivity, and food webs. Collectively, these areas produced 83% of global fisheries yields over the past 50 years. Fish diversity data for each LME were derived independently from a comprehensive fish taxonomic database (*22*).

Globally, the rate of fisheries collapses, defined here as catches dropping below 10% of the recorded maximum (*23*), has been accelerating over time, with 29% of currently fished species considered collapsed in 2003 (Fig. 3A, diamonds). This accelerating trend is best described by a power relation ($y = 0.0168x^{1.8992}$, $r = 0.96$, $P < 0.0001$), which predicts the percentage of currently collapsed taxa as a function of years elapsed since 1950. Cumulative collapses (including recovered species) amounted to 65% of recorded taxa (Fig. 3A, triangles; regression fit: $y = 0.0227x^{2.0035}$,

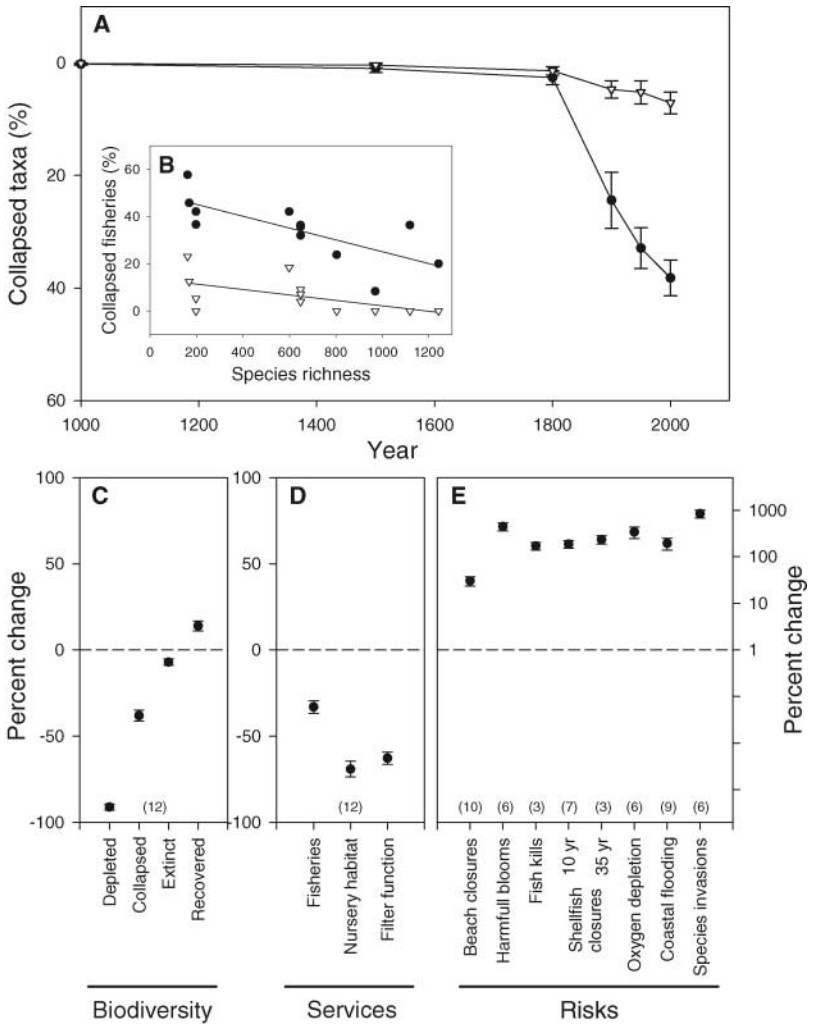


Fig. 2. Regional loss of species diversity and ecosystem services in coastal oceans. (**A**) Trends of collapse (circles, >90% decline) and extinction (triangles, 100% decline) of species over the past 1000 years. Means and standard errors are shown ($n = 12$ regions in Europe, North America, and Australia). (**B**) Percentage of collapsed (circles) and extinct (triangles) fisheries in relation to regional fish species richness. Significant linear regression lines are depicted ($P < 0.01$). (**C** to **E**) Relative losses or gains in (C) biodiversity, (D) ecosystem services, and (E) risks that are associated with the loss of services. The number of studies is given in parentheses; error bars indicate standard errors.

$r = 0.96$, $P < 0.0001$). The data further revealed that despite large increases in global fishing effort, cumulative yields across all species and LMEs had declined by 13% (or 10.6 million metric tons) since passing a maximum in 1994.

Consistent with the results from estuaries and coastal seas (Fig. 2B), we observed that these collapses of LME fisheries occurred at a higher rate in species-poor ecosystems, as compared with species-rich ones (Fig. 3A). Fish diversity varied widely across LMEs, ranging from ~20 to 4000 species (Fig. 3B), and influenced fishery-related services in several ways. First, the proportion of collapsed fisheries decayed exponentially with increasing species richness (Fig. 3C). Furthermore, the average catches of non-collapsed fisheries were higher in species-rich systems (Fig. 3D). Diversity also seemed to increase robustness to overexploitation. Rates of recovery, here defined as any post-collapse increase above the 10% threshold, were positively correlated with fish diversity (Fig. 3E). This positive relationship between diversity and recovery became stronger with time after a collapse (5 years, $r = 0.10$; 10 years, $r = 0.39$; 15 years, $r = 0.48$). Higher taxonomic units (genus and family) produced very similar relationships as species richness in Fig. 3; typically, relationships became stronger with increased taxonomic aggregation. This may suggest that taxonomically related species play complementary functional roles in supporting fisheries productivity and recovery.

A mechanism that may explain enhanced recovery at high diversity is that fishers can switch more readily among target species, potentially providing overfished taxa with a chance to recover. Indeed, the number of fished taxa was a log-linear function of species richness (Fig. 3F). Fished taxa richness was negatively related to the variation in catch from year to year (Fig. 3G) and positively correlated with the total production of catch per year (Fig. 3H). This increased stability and productivity are likely due to the portfolio effect (*24*, *25*), whereby a more diverse array of species provides a larger number of ecological functions and economic opportunities, leading to a more stable trajectory and better performance over time. This portfolio effect has independently been confirmed by economic studies of multispecies harvesting relationships in marine ecosystems (*26*, *27*). Linear (or log-linear) relationships indicate steady increases in services up to the highest levels of biodiversity. This means that proportional species losses are predicted to have similar effects at low and high levels of native biodiversity.

Marine reserves and fishery closures. A pressing question for management is whether the loss of services can be reversed, once it has occurred. To address this question, we analyzed available data from 44 fully protected marine reserves and four large-scale fisheries closures (*15*). Reserves and closures have been used to reverse the decline of marine biodiversity on local and regional scales (*28*, *29*). As such, they can be viewed as replicated large-scale experiments. We used meta-analytic techniques (*15*) to test for consistent trends in biodiversity and services across all studies (Fig. 4).

We found that reserves and fisheries closures showed increased species diversity of target and nontarget species, averaging a 23% increase in species richness (Fig. 4A). These increases in biodiversity were associated with large increases in fisheries productivity, as seen in the

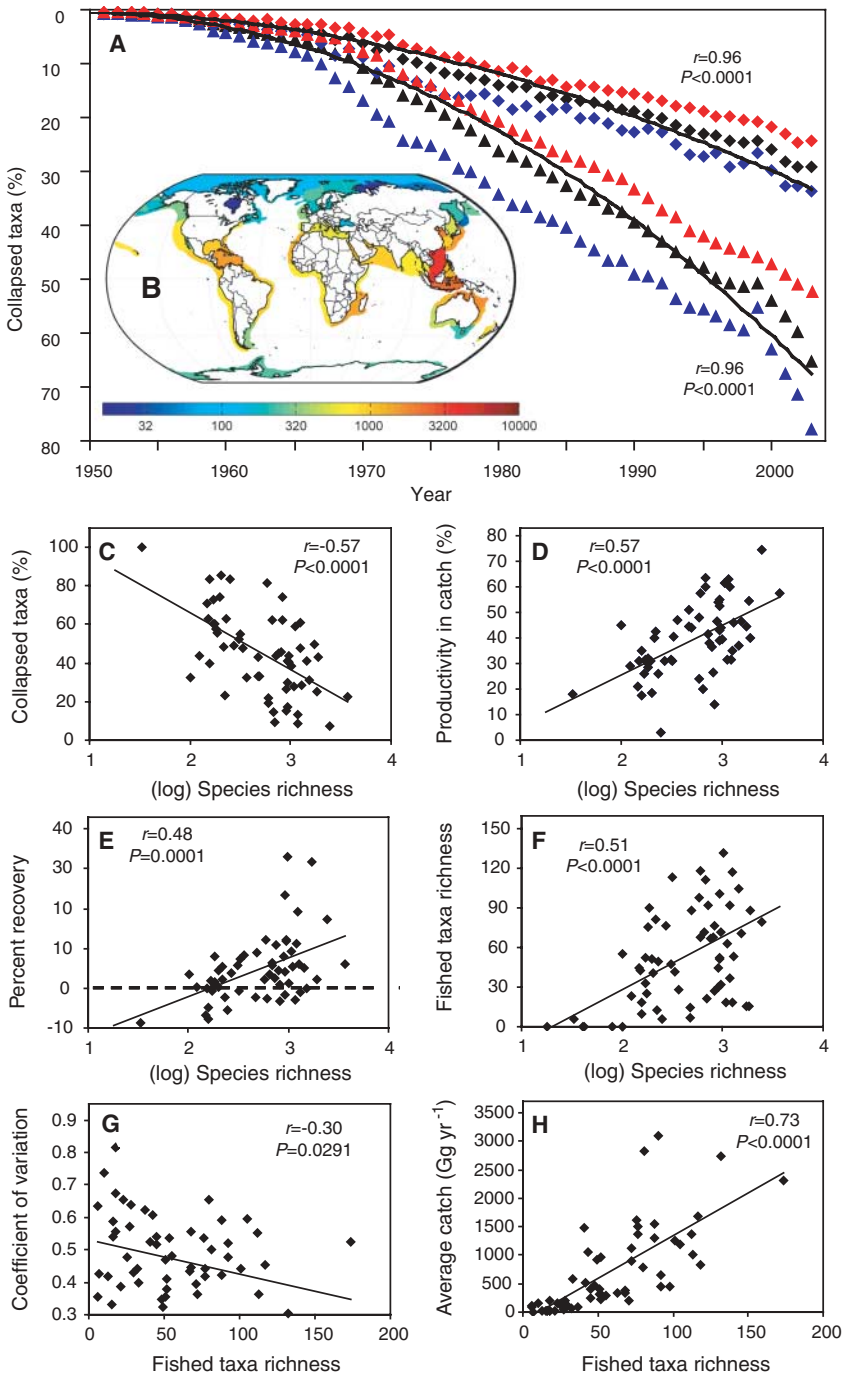


Fig. 3. Global loss of species from LMEs. (**A**) Trajectories of collapsed fish and invertebrate taxa over the past 50 years (diamonds, collapses by year; triangles, cumulative collapses). Data are shown for all (black), species-poor (<500 species, blue), and species-rich (>500 species, red) LMEs. Regression lines are best-fit power models corrected for temporal autocorrelation. (**B**) Map of all 64 LMEs, color-coded according to their total fish species richness. (**C**) Proportion of collapsed fish and invertebrate taxa, (**D**) average productivity of noncollapsed taxa (in percent of maximum catch), and (**E**) average recovery of catches (in percent of maximum catch) 15 years after a collapse in relation to LME total fish species richness. (**F**) Number of fished taxa as a function of total species richness. (**G**) Coefficient of variation in total catch and (**H**) total catch per year as a function of the number of fished taxa per LME.

fourfold average increase in catch per unit of effort in fished areas around the reserves (Fig. 4B). The difference in total catches was less pronounced (Fig. 4B), probably because of restrictions on fishing effort around many reserves. Resistance and recovery after natural disturbances from storms and thermal stress tended to increase in reserves, though not significantly in most cases (Fig. 4C). Community variability, as measured by the coefficient of variation in aggregate fish biomass, was reduced by 21% on average (Fig. 4C). Finally, tourism revenue measured as the relative increase in dive trips within 138 Caribbean protected areas strongly increased after they were established (Fig. 4D). For several variables, statistical significance depended on how studies were weighted (Fig. 4, solid versus open circles). This is probably the result of large variation in sample sizes among studies (*15*). Despite the inherent variability, these results suggest that at this point it is still possible to recover lost biodiversity, at least on local to regional scales; and that such recovery is generally accompanied by increased productivity and decreased variability, which translates into extractive (fish catches around reserves) and nonextractive (tourism within reserves) revenue.

Conclusions. Positive relationships between diversity and ecosystem functions and services were found using experimental (Fig. 1) and correlative approaches along trajectories of diversity loss (Figs. 2 and 3) and recovery (Fig. 4). Our data highlight the societal consequences of an ongoing erosion of diversity that appears to be accelerating on a global scale (Fig. 3A). This trend is of serious concern because it projects the global collapse of all taxa currently fished by the mid–21st century (based on the extrapolation of regression in Fig. 3A to 100% in the year 2048). Our findings further suggest that the elimination of locally adapted populations and species not only impairs the ability of marine ecosystems to feed a growing human population but also sabotages their stability and recovery potential in a rapidly changing marine environment.

We recognize limitations in each of our data sources, particularly the inherent problem of inferring causality from correlation in the larger-scale studies. The strength of these results rests on the consistent agreement of theory, experiments, and observations across widely different scales and ecosystems. Our analysis may provide a wider context for the interpretation of local biodiversity experiments that produced diverging and controversial outcomes (*1*, *3*, *24*). It suggests that very general patterns emerge on progressively larger scales. High-diversity systems consistently provided more services with less variability, which has economic and policy implications. First, there is no dichotomy between biodiversity conservation and long-term economic development; they must be viewed as interdependent societal goals. Second, there was no evidence for redundancy at high levels of diversity; the improvement of services was continuous on a log-linear scale (Fig. 3). Third, the buffering impact of species diversity on the resistance and recovery of ecosystem services generates insurance value that must be incorporated into future economic valuations and management decisions. By restoring marine biodiversity through sustainable fisheries management, pollution control, maintenance of essential habitats, and the creation of marine reserves, we can invest in the productivity and reliability of the goods and services that the ocean provides to humanity. Our analyses suggest that business as usual would foreshadow serious threats to global food security, coastal water quality, and ecosystem stability, affecting current and future generations.

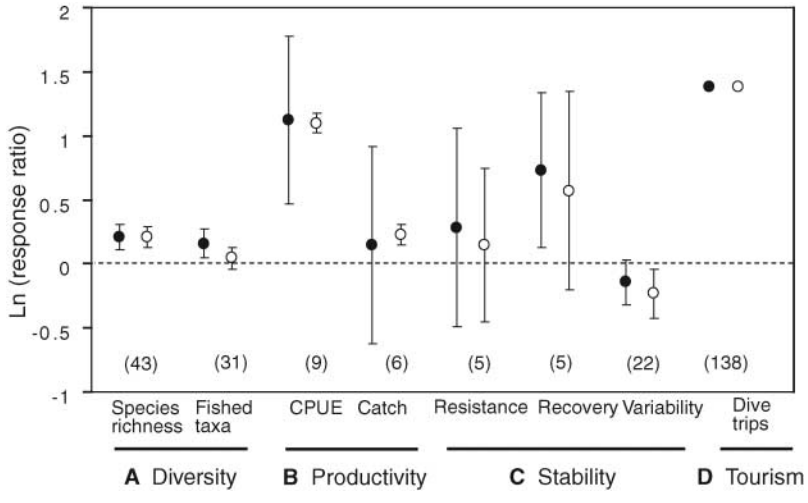


Fig. 4. Recovery of diversity and ecosystem services in marine protected areas and fisheries closures. Shown are the response ratios (inside versus outside the reserve or before and after protection ±95% CI) of (**A**) species diversity and (**B** to **D**) ecosystem services that correspond to fisheries productivity, ecosystem stability, and tourism revenue, respectively. Positive values identify increases in the reserve relative to the control; error bars not intersecting zero indicate statistical significance ($P < 0.05$). Solid circles represent unweighted averages; open circles are weighted by sample size (see supporting online methods for details). The number of studies is shown in parentheses. CPUE, catch per unit of effort.

References and Notes

1. M. Loreau *et al.*, *Science* **294**, 804 (2001).
2. M. Palmer *et al.*, *Science* **304**, 1251 (2004).
3. D. U. Hooper *et al.*, *Ecol. Monogr.* **75**, 3 (2005).
4. I. E. Hendriks, C. M. Duarte, C. H. R. Heip, *Science* **312**, 1715 (2006).
5. C. H. Peterson, J. Lubchenco, in *Nature's Services: Societal Dependence on Natural Ecosystems*, G. C. Daily, Ed. (Island Press, Washington, DC, 1997), pp. 177–194.
6. C. M. Holmlund, M. Hammer, *Ecol. Econ.* **29**, 253 (1999).
7. F. Danielsen *et al.*, *Science* **310**, 643 (2005).
8. W. N. Adger, T. P. Hughes, C. Folke, S. R. Carpenter, J. Rockstrom, *Science* **309**, 1036 (2005).
9. N. K. Dulvy, Y. Sadovy, J. D. Reynolds, *Fish Fish.* **4**, 25 (2003).
10. H. K. Lotze *et al.*, *Science* **312**, 1806 (2006).
11. J. M. Pandolfi *et al.*, *Science* **301**, 955 (2003).
12. J. B. C. Jackson *et al.*, *Science* **293**, 629 (2001).
13. B. Worm, M. Sandow, A. Oschlies, H. K. Lotze, R. A. Myers, *Science* **309**, 1365 (2005).
14. D. Raffaelli, *Science* **306**, 1141 (2004).
15. Details on methods and data sources are available as supporting material on *Science* Online.
16. A. R. Hughes, J. J. Stachowicz, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8998 (2004).
17. T. B. H. Reusch, A. Ehlers, A. Hämmerli, B. Worm, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2826 (2005).
18. R. Dame *et al.*, *Aquat. Ecol.* **36**, 51 (2002).
19. J. J. Stachowicz, R. B. Whitlatch, R. W. Osman, *Science* **286**, 1577 (1999).
20. R. Watson, A. Kitchingman, A. Gelchu, D. Pauly, *Fish Fish.* **5**, 168 (2004).
21. K. Sherman, A. Duda, *Mar. Ecol. Prog. Ser.* **190**, 271 (1999).
22. R. Froese, D. Pauly, Eds., *FishBase* (www.fishbase.org, version 12/2004).
23. R. Froese, K. Kesner-Reyes, *Impact of Fishing on the Abundance of Marine Species* [ICES Council Meeting Report CM 12/L:12, International Council for the Exploration of the Sea (ICES), Copenhagen, Denmark, 2002].
24. D. Tilman, *Ecology* **80**, 1455 (1999).
25. D. Tilman, P. B. Reich, J. M. H. Knops, *Nature* **441**, 629 (2006).
26. H. Wacker, *Res. Energy Econ.* **21**, 89 (1999).
27. D. Finnoff, J. Tschirhart, *J. Environ. Econ. Manage.* **45**, 589 (2003).
28. C. M. Roberts, J. P. Hawkins, *Fully-Protected Marine Reserves: A Guide* (World Wildlife Fund, Washington, DC, 2000), pp. 241–246.
29. S. R. Palumbi, in *Marine Community Ecology*, M. D. Bertness, S. D. Gaines, M. E. Hay, Eds. (Sinauer, Sunderland, MA, 2001), pp. 510–530.

30. This work was conducted as part of the Linking Marine Biodiversity and Ecosystem Services Working Group, supported by the National Center for Ecological Analysis and Synthesis funded by NSF, the University of California, and the Santa Barbara campus. The project was stimulated by N. Loder after discussion at the conference Marine Biodiversity: The Known, Unknown, and Unknowable, funded by the Sloan Foundation. The authors thank D. Pauly and the Sea Around Us Project (http://seaaroundus.org), supported by the the Pew Charitable Trusts, for access to global catch data; W. Blanchard and M. Sandow for technical support; E. Green for dive trip data; and N. Baron, P. Kareiva, R. A. Myers, U. Sommer, and D. Tittensor for helpful comments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/787/DC1
Methods and Data Sources
Tables S1 to S5
References

10 July 2006; accepted 3 October 2006
10.1126/science.1132294

CORRECTED 22 DECEMBER 2006; SEE LAST PAGE

Giant Ringlike Radio Structures Around Galaxy Cluster Abell 3376

Joydeep Bagchi,[1]* Florence Durret,[2] Gastão B. Lima Neto,[3] Surajit Paul[4]

In the current paradigm of cold dark matter cosmology, large-scale structures are assembling through hierarchical clustering of matter. In this process, an important role is played by megaparsec (Mpc)–scale cosmic shock waves, arising in gravity-driven supersonic flows of intergalactic matter onto dark matter–dominated collapsing structures such as pancakes, filaments, and clusters of galaxies. Here, we report Very Large Array telescope observations of giant (~2 Mpc by 1.6 Mpc), ring-shaped nonthermal radio-emitting structures, found at the outskirts of the rich cluster of galaxies Abell 3376. These structures may trace the elusive shock waves of cosmological large-scale matter flows, which are energetic enough to power them. These radio sources may also be the acceleration sites where magnetic shocks are possibly boosting cosmic-ray particles with energies of up to 10^{18} to 10^{19} electron volts.

A large fraction (~30%) of the baryon mass of the universe at the present epoch resides in a tenuous gas phase known as the warm-hot intergalactic medium (WHIM) (*1*). Although the WHIM is heated to ≈ 10^5 to 10^7 K by intergalactic shocks, its direct detection is very challenging owing to its low density and temperature. The main tracers of this gas are highly excited oxygen lines mainly visible in soft x-rays and the far-ultraviolet, whose emission or absorption signal is well below the sensitivity threshold of the current instruments. Alternatively, the supersonic infalls and resulting shock waves propagating in the magnetized intergalactic medium (IGM) or WHIM around galaxy clusters and filaments can be probed by the synchrotron and inverse-Compton radiation emitted by energetic electrons accelerated at the shock fronts (*2*, *3*). For particles to be shock accelerated, magnetic fields need to be present at the shocks for generation of Alfvén waves (transverse waves moving along magnetic-field lines embedded in an electrically conducting fluid) and to-and-fro scattering of particles across the shock front by wave-particle interactions [the so called diffusive shock acceleration process (*4*)]. The origin of cosmic magnetic fields is currently unknown (*5*), and so any observational evidence for them in the IGM environment is of great importance. Magnetic fields (~μG) are observed in the denser, and hotter (relative to WHIM) intracluster medium (ICM) of clusters of galaxies by means of the Faraday effect, causing rotation of the plane of polarization of light of the background radio sources (*6*). Magnetic fields are further revealed by the presence of clusterwide Mpc-scale (1 Mpc = 3.08×10^{24} cm) radio-halos and peripheral radio-relics (*7*, *8*) (both are large-scale diffuse radio sources, having no optical counterparts and no obvious connection to the galaxies in clusters) in a few clusters of galaxies. Radio-halos and radio-relics are believed to be the result of synchrotron emission of relativistic electrons shock accelerated in cluster mergers (*9*).

We have discovered with the Very Large Array (VLA) telescope, at 1.4-GHz frequency, a large-scale [~2 h^{-1} Mpc (*10*)] ringlike synchrotron radio-emission structure, possibly tracing the intergalactic shocks around a rich cluster of galaxies Abell 3376, at redshift z = 0.046 (*11*). The southern galaxy cluster Abell 3376 was observed with the VLA, a Y-shaped interferometric array of 30 25-m-diameter radio telescopes located on the plains of San Agustin in Socorro, New Mexico, the United States (*12*). Figure 1 shows the most important findings: a pair of giant, "arc" shaped diffuse radio-emitting sources, each with a linear dimension ~1 h^{-1} Mpc, located at the outskirts of this cluster at the projected distance of ~1 h^{-1} Mpc from the cluster center. In addition, the eastern structure shows several more thin filamentary structures behind the radio arc and a peculiar looplike feature joining two linear filaments (Fig. 1). This radio image is the result of combining data in the VLA CnB and DnC hybrid configurations, having the sensitivity for mapping diffuse, large-scale emissions, as well as the angular resolution (~20 arc sec) for identifying the superposed background point sources (*12*). The VLA radio image shown has ~40 microjansky (μJy) per beam of root mean square noise background, and the estimated signal-to-noise ratio for detection of these giant radio arcs ranges from ~3σ at the faintest contour to ~25σ at the peaks (after removing a few superposed background point sources, which are visible in Fig. 1). Noticeably, both radio structures are positioned with their concave side facing the cluster center, and they fit quite well on the surface of a large, projected elliptical ringlike formation of dimension ~2 Mpc by 1.6 Mpc, oriented in position angle ≈ 85°. The center of this ellipse falls at the position right ascension ≈ $06^h\ 01^m\ 32^s$, and declination ≈ −39°59′50″, which is taken as the center of symmetry of this cluster (marked "+" in Fig. 1). No other "radio-halo" type of diffuse, large-scale emission was found near the cluster center down to the surface brightness limit mentioned above.

The luminous thermal bremsstrahlung x-ray emission detected by the Position Sensitive Proportional Counter (PSPC) detector onboard the Roentgen Satellite (ROSAT) is shown in Fig. 1 (in the 0.14- to 2.0-keV band). In the central region, it reveals a highly disturbed, nonequilibrium state of the intracluster thermal gas, which is obviously extended like a "comet" or "bullet-head," its wake extending along the major axis of the elliptical radio structure described above. Both radio sources are located on the outer periphery of the x-ray emission observed by ROSAT, which has an unblocked field of view of 17.5 arc min radius. From galaxy redshift measurements in Abell 3376, a filamentary structure composed of at least three major galaxy groups is found, oriented along the general elongation of x-ray emission mapped by ROSAT and X-ray Multi Mirror Mission (XMM)–Newton (*13*, *14*).

XMM-Newton and ROSAT x-ray observations (*12*) shown in Fig. 2 reveal strong evidence for merger activity of subclusters along this filamentary axis. The field of view of XMM is a circle of radius 15 arc min, and the image shown is in the 0.3- to 8.0-keV band. The XMM map of Fig. 2 is quite similar to the ROSAT image in Fig. 1, but it shows more clearly the "bullet-head" and multiple x-ray peaks to the south southwest of it, each one probably associated with a merging group. The x-ray temperature map (Fig. 2) has typical errors of about 10%. This map reveals an overall temperature of about 5 keV (5.8×10^7 K), with several alternating hot and cold regions crossing the cluster, divided by a prominent "cold-arc" at about 3 keV, which originates at the north edge and curves southward toward the east. The most plausible scenario is that a large group or a small cluster is falling onto the main body of the cluster from the east-northeast, thus creating hotter regions through a shock. Such a scenario would agree with the second-brightest galaxy on the x-ray peak associated with the strong radio source MRC 0600-399, of which both radio-jets are bent backward toward east-northeast, away from the "bullet," suggesting an infall and ram pressure on radio jets from a gaseous "wind" blowing from the west-southwest direction along position angle (PA) ~ 70° (Fig. 1; detailed image in fig. S1). The dominant elliptical galaxy in Abell 3376 is located in the south-west subcluster, ~18 arc min (~1 Mpc) from the x-ray peak (Fig. 2). The highly asymmetrical metal distribution [mainly iron, about 30% errors (*12*)] near the center also suggests a violent and recent dynamical event, where the

[1]The Inter-University Centre for Astronomy and Astrophysics (IUCAA), Post Bag 4, Ganeshkhind, Pune University Campus, Pune 411007, India. [2]Institut d'Astrophysique de Paris, UMR-7095, CNRS, Université Pierre et Marie Curie, Paris, France, and Observatoire de Paris-Meudon, LERMA, Paris, France. [3]Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, Brazil 05508-900. [4]Institut fur Theoretische Physik und Astrophysik, Universität Würzburg, Sanderring 2, 97070 Würzburg, Germany.

*To whom correspondence should be addressed. E-mail: joydeep@iucaa.ernet.in

gas is not well mixed and we observe patches of high and low metallicity (Fig. 2; metallicity profile in fig. S2).

What is the energy source that could possibly power such giant radio structures? Any electron acceleration process must be at least energetic enough to account for $\nu I_\nu \approx 2.1 \times 10^{40}$ erg/s of radio emission observed at the frequency $\nu = 1.4$ GHz (sum of radio luminosity of two sides). First, there is no evidence that any optical galaxy is obviously associated with the radio arcs (Fig. 1), and it is unlikely that these arcs are the usual cluster radio galaxies. On the other hand, these radio structures could be generic to radio-relics known to occur in certain merging clusters (*7*, *8*), although such double relics are rare. Another very similar configuration of Mpc-scale binary radio-relic arcs is found in the well-known merging galaxy cluster Abell 3667 (*15*). Second, initial acceleration at a central active galactic nucleus–like point source and particle transport by diffusion across a ~Mpc scale is not possible. The radiative energy loss time scale t_{IC} for an electron of relativistic Lorentz factor γ is $t_{IC} \approx 2.3 \times 10^8(\gamma/10^4)^{-1}\,(1+z)^{-4}$ year [assuming only inverse Compton scattering (IC) on the ambient 2.7 K background and a weak magnetic field $B \ll 3$ μG], and the diffusion length within the IC cooling time: $l_{diff} \approx (D_B\, t_{IC})^{1/2} = 11.36\,(B/\mu G)^{-1/2}$ pc $\ll 1$ Mpc, the scale of the observed radio arcs [here $D_B = 1.7 \times 10^{23}(\gamma/10^4)\,(B/\mu G)^{-1}$ cm^2/s, the Bohm diffusion coefficient, for scattering on saturated field fluctuations]. The discrepancy between the Bohm diffusion length-scale and the radio structure size is so large that, even with the inclusion of advective transport by bulk flows and more effective diffusion in ordered magnetic fields, electrons are still unable to cross the emission region within a radiative lifetime. Electrons must be accelerated to $\gamma \approx 1.8 \times 10^4$ (~10 GeV) for synchrotron emission at frequency $\nu_{syn} = 1.4$ GHz in a magnetic field $B = 1$ μG ($\nu_{syn} \propto \gamma^2 B$). Sparse observations (*2*) and simulations (*16*) of magnetic fields in the diffuse IGM at the cluster outskirts and within filaments suggest that $B \sim 10^{-7}$ to 10^{-9} G. In this situation, the synchrotron loss time scale is $t_{syn} \approx 10^3(B/0.1\ \mu G)^{-2}\,(1+z)^4\, t_{IC}$, and IC radiation is the dominant cooling process.

Clearly, the detection of these large radio structures at ~Mpc distances from the cluster center inevitably requires some form of in situ acceleration mechanism for particles and the magnetic field powering them. The detected structures cannot result from the chance superposition of background radio point sources, which are very few (~0.02 sources per arc min^2 above 1-mJy flux density), and most of the emission is truly diffuse. Lastly, it is very unlikely that these diffuse sources are the radio-lobes of a currently active giant radio galaxy (GRG), because up to ≈0.2 mJy/beam surface brightness limit (≈5σ signal) no radio jets or plumes connect the radio sources with any central optical galaxy, and GRGs of such extreme size (~2 Mpc) are not commonly found (*17*). On the other hand, their concave bow-shock–like structure, symmetric and tangential juxtaposition on the merger axis—tangential both to the chain of subclusters of galaxies and to the x-ray emission elongation axis (Fig. 1)—strongly supports association of these arcs with the cluster and suggests their origin in a cosmological-scale energetic event linked to the cluster-formation process.

Hydrodynamical simulations show that only shock waves induced in structure-formation processes are sufficiently extended, long-lived, and energetic to overcome the strong radiation losses of the relativistic electrons over a Mpc scale, rapidly accelerating them to relativistic energies (*18*, *19*), which could emit synchrotron radio emission at the level observed (*3*). These ubiquitous shocks are an inevitable consequence of gravitational collapse, and they are pivotal to virialization of diffuse IGM. Two competing shockwave-inducing mechanisms are plausible: (i) Outgoing "merger shocks" emanating from the cluster center, induced by mergers of subclusters; and (ii) the accretion flows of IGM and associated "accretion shocks" near the virial radius. We briefly mention both these plausible models because their observational signatures—which cannot be discerned in the present data—are very similar.

In the merger shock model, a pair of outgoing, merger-generated shock waves (*20*) could have created these diffuse radio sources on the outskirts of Abell 3376 by compressing and accelerating a preexisting lower energy "fossil" electron population (*21*). It takes ~10^9 years for shock fronts to cross the ~1-Mpc distance from the cluster center (at the "+" mark in Fig. 1) if the shock speed is $V_s = 10^3$ km/s. The merger shock model adequately explains the extended steep-spectrum radio arcs and elongated x-ray structure found in a well-studied cluster, Abell 3667 (*22*). This is a plausible model given the striking resemblance of Abell 3376 and Abell 3667, both of which provide strong evidence for

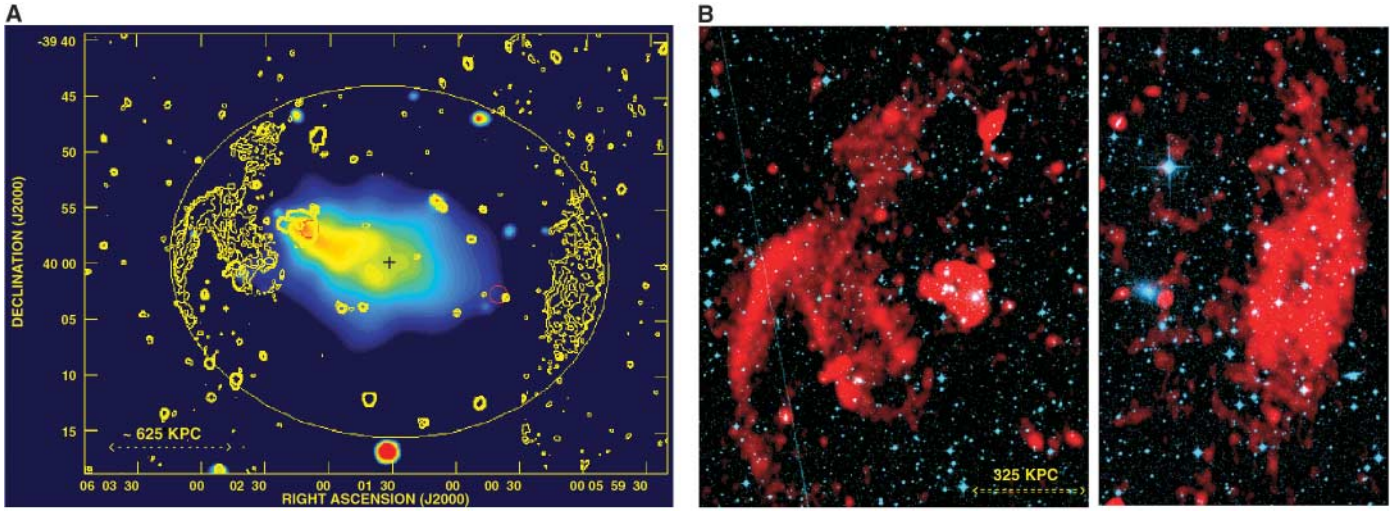


Fig. 1. (**A**) A composite map of radio and x-ray emissions from the galaxy cluster Abell 3376. The radio emission is represented by yellow contours (0.12, 0.24, 0.48, and 1 mJy per beam; beam width: 20 arc sec full width at half maximum Gaussian) obtained from the VLA 1.4-GHz observations (*12*). The yellow ellipse shows an elliptical fit to the peripheral radio structures, and the "+" marks the center of the ellipse. The central color image depicts the thermal bremsstrahlung x-ray emission detected by the Position Sensitive Proportional Counter instrument onboard the Roentgen Satellite (≈12-ks exposure, within 0.14- to 2.0-keV band). The red circles mark the position of the two brightest cluster galaxies—the brightest elliptical galaxy on the lower right and the second brightest elliptical galaxy associated with the bent-jet radio source MRC 0600-399 near the x-ray peak. (**B**) Composite images obtained from superposing the radio and optical images. The VLA 1.4-GHz radio maps (in red) for the eastern (left) and the western (right) radio structures are shown overlayed on the red band Digitized Sky Survey image (in blue).

mergers and have similar x-ray and radio morphologies (*15*). The peripheral location of quasi-ringlike radio arcs and the absence of a central radio halo in Abell 3376 agree with simulations of merger shocks (*21*) (fig. S3).

In the accretion model, these giant radio structures could be tracing the accretion shocks from infall of IGM at the outskirts of the cluster Abell 3376. An accretion shock at the virial radius propagates outward while infalling gas crosses it inward supersonically, which could lead to shock acceleration of particles and radio emission if the infalling gas is magnetized [as in (*9*)]. The virial radius is $r_{vir} = \{3M/[4\pi\Delta_c(z)\rho_{crit}(z)]\}^{1/3}$, where $\Delta_c(z)$ is the ratio of mass density to critical density $\rho_{crit}(z)$ (*10*) inside a dissipationless, spherical, collapsing virializing halo of mass M at redshift z. In Λ cold dark matter and Einstein–de Sitter (*10*) cosmologies, $\Delta_c \approx 340$ and $\Delta_c \approx 178$, respectively, at $z = 0$ (*23*), implying that r_{vir} = 1.4 to 1.7 Mpc for Abell 3376 and the cluster virial mass is $M_{vir} = 5.2 \times 10^{14}$ solar masses ($M_\odot$) (*24*). This exceeds the observed ~1-Mpc distance at which radio sources are found, possibly due to a nonspherical filamentary geometry (Fig. 1) and an apparent linear foreshortening due to projection on the observer's sky plane.

Recent hydrodynamical simulations reveal an intergalactic shock structure more complex than assumed in a simple spherical accretion model described above. Around rich clusters these shocks (both accretion and merger) do not form fully illuminated rings (in projection), but they are strongest along the axes where filaments funnel the IGM deep into the forming clusters (*18*, *19*). Along the same directions, high-speed mergers of subclusters (with merger shocks) may take place, as observed here and in cluster Abell 3667 (*15*). This may provide a clue to the origin of the radio structures observed in Abell 3376, which are located on an elliptical ringlike formation in projection (Fig. 1), yet only sections of this ring are actually illuminated along an axis, where x-ray data clearly show that mergers and infalls are taking place and where shocks are strongest, leading to more efficient particle acceleration. The actual three dimensional morphology could be like a `"bubble" or "egg-shell" on the surface of which the shocks are located within a narrow zone. Some recent numerical simulations of structure shocks around similar massive galaxy clusters (10^{14} to 10^{15} $M_\odot$) do predict this quasi-ringlike geometry of nonthermal radiation (*19*, *21*).

Therefore, in the giant radio structures discovered in Abell 3376 (and Abell 3667), we are probably witnessing the first observational evidence for merger or accretion (or even both) shocks near the sparsely studied virial infall region of a massive galaxy cluster. These shocks are an important probe of the gas dynamics at the transition zone between the hot cluster medium and cold WHIM gas. The Mach numbers (M_s) associated with these shock structures are not known as yet. Their estimation requires temperature data on both sides of the shock to

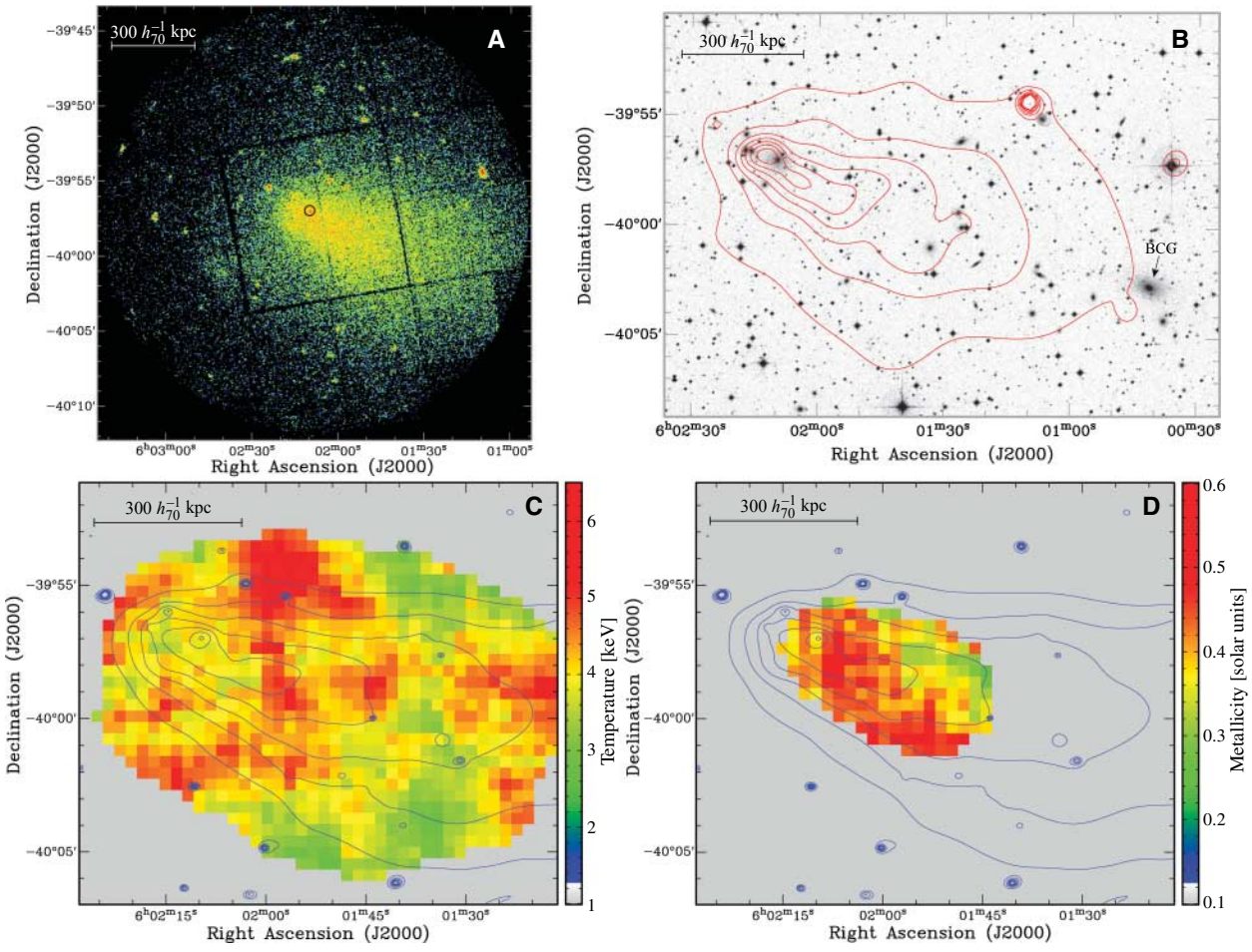


Fig. 2. (**A**) XMM-Newton MOS1 and MOS2 [MOS stands for Metal Oxide Semiconductor (*12*)] telescope composite photon image in the 0.3- to 8.0-keV band. The black circle shows the position of the second-brightest galaxy, which coincides with the x-ray peak. (**B**) Red band Digitized Sky Survey optical image with ROSAT smoothed x-ray intensity iso-contours superposed. The ROSAT energy range is 0.2 to 2.0 keV. The brightest cluster galaxy (BCG) is indicated with an arrow. The BCG sits at the edge of the ROSAT field of view but outside of the XMM field of view. (**C**) Temperature map (color scale in keV units) derived from XMM data. Noticeably, there are several alternating cold and hot regions, their temperature varying from ~2 to ~6 keV. The superposed x-ray intensity contours are from an adaptively smoothed XMM image. (**D**) Metallicity map (color scale in solar units) derived from XMM data (*12*). It shows a strong metallicity variation along the x-ray intensity elongation axis. Contours are the same as in the previous panel. The scale bars are in kpc, assuming redshift z = 0.046 and standard Λ cold dark matter cosmology (*10*).

fit Rankine-Hugoniot equations. For an intergalactic shock at the cluster outskirts in cold medium, simulations predict high Mach numbers [$M_s >> 1$ (*25*)], and the pre- and postshock temperatures T_1 and T_2 are related as $T_2/T_1 \approx (5/16) M_s^2$ in a gas of specific heat ratio $\Gamma = 5/3$. Setting $T_2 \sim T_{vir} = 4 \times 10^7$ K, the virial temperature (T_{vir}) for the virial mass of the cluster, the temperature of the infalling gas spans the range ~5×10^4 to 5×10^6 K, for M_s between 50 and 5. This is close to the expected temperature range of WHIM gas (*1*), although too cold to be detectable in a keV-band x-ray observation (Fig. 1).

The exact physical mechanism of the electron acceleration is yet to be revealed, but diffusive shock acceleration (DSA) is a strong possibility, which explains the ~4 Mpc–scale diffuse radio emission from the filamentary protocluster ZwCl 2341.1+0000 (*2*). In DSA, the time scale for a particle (electron, proton, or a heavier ion) to reach energy E is (*4*) $t_{acc}(E) \approx 8\, D_B/u_3^2 = 8.45 \times 10^9\, u_3^{-2}\, E_{19}\, B_\mu^{-1} Z^{-1}$ year (for a strong shock of $M_s >> 1$, and Bohm diffusion). Here, $E_{19} = (E/10^{19}$ eV), $u_3 = (V_{shock}/10^3$ km s^{-1}), $B_\mu = (B/10^{-6}$ G), and Z is the ionic charge ($Z = 1$ for electron or proton). For cosmic-ray protons, which suffer negligible radiative losses below 50×10^{19} eV, the highest acceleration energy, E^p_{max}, is limited only by the finite lifetime of shocks, i.e., $t_{acc} = t_{merger} \sim 10^9$ to 10^{10} year (here t_{merger} is the time scale for merger), thus giving $E^p_{max} \sim 10^{18}$ to 10^{19} eV. Heavier ions carrying more charge (for example, ^{26}Fe) could be accelerated up to much larger energy, limited by loss processes. For cosmic-ray electrons, substantial radiative losses would limit their energy to a maximum $E^e_{max} \sim 3.73 \times 10^{13}\, u_3\, B_\mu^{-1/2}$ eV. Lower-energy electrons of ~$10\, B_\mu^{-1/2}$ GeV would radiate the observed 1.4-GHz synchrotron radio emission. The sketched scenario is consistent with hydrodynamical simulations (*26*), which show that during gravitational infall and violent mergers in clusters, protons can be accelerated up to the energy $E^p_{max} \sim 5 \times 10^{19}$ eV, if a turbulent magnetic field of ~0.1 to 1 μG is available and if a fraction about 10^{-4} of the infalling kinetic energy can be injected into the IGM as high-energy particles. Our estimate of the minimum-energy magnetic field (B_{min}) within the radio-arc regions in Abell 3376 is $B_{min} = 0.5$ to 3.0 μG, which depends on model parameters (*27*).

Our observations here represent a substantial advance in the field, because they probe several important components of the cosmic environment: intergalactic gas, magnetic fields, and cosmic-rays. They indicate that magnetic fields of appreciable strength are present not only in the ICM but also in the diffuse intergalactic medium, i.e., in the gas that will be shocked as it accretes onto collapsing structures—the precursors of virialized galaxy clusters. These magnetic fields are necessary for providing the scattering centers for the diffusive shock acceleration mechanism, and also for the synchrotron emission that we have observed. Because it is not obvious how magnetic fields are amplified up to such large values along filaments, our observation poses further challenges to theoretical models (*28*, *29*).

On the other hand, as we show in this work, if diffusive shock acceleration takes place with some efficiency during the nonlinear stage of large-scale structure formation, cosmic-ray ions accumulating in the forming structure could become dynamically important (*30*). Objects like Abell 3376 could be acceleration sites for cosmic-rays, where magnetized shocks are possibly boosting hadronic particles to relativistic energies up to 10^{18} to 10^{19} eV. Downstream from the shocks (toward the cluster center), accelerated protons will be transported by diffusion and advection, over a length scale (l_{diff}) comparable to the cluster size, $l_{diff} \approx 1.08\, u_3^{-1}\, E_{19}\, B_\mu^{-1}$ Mpc, under the Bohm diffusion regime (*25*). These protons will remain trapped within the cluster volume for a time scale comparable to the age of the cluster until their energy approaches $E^p \gtrsim 2 \times 10^{17}\, B_\mu$ eV. Inverse Compton scatter of the 2.7 K cosmic microwave background photons from both primary electrons accelerated in shocks and from secondary electrons originating in hadronic processes (mainly pion decay and pair production) will give rise to photons of energies $E_\gamma \sim 100\, \gamma_7^{e2}$ GeV (where γ_7^e is the electron Lorentz factor in units of 10^7). This inverse Compton spectrum could extend up to ~10 TeV, limited by a likely high-energy cut-off in the electron energy spectrum. Radiation from primary electrons would trace the current shock locations, owing to their short radiative cooling time. This tell-tale radiation might be detectable with hard x-ray (SUZAKU, NEXT), or 1 MeV–100 GeV (GLAST), or high-energy TeV γ-ray (CANGAROO, MAGIC, VERITAS, HESS) telescopes, opening a new window on the nonthermal processes in cosmological large-scale flows.

References and Notes

1. R. Cen, J. P. Ostriker, *Astrophys. J.* **519**, L109 (1999).
2. J. Bagchi *et al.*, *New Astron.* **7**, 249 (2002).
3. U. Keshet, E. Waxman, A. Loeb, *Astrophys. J.* **617**, 281 (2004).
4. L. O'C. Drury, *Rep. Prog. Phys.* **46**, 973 (1983).
5. C. L. Carilli, G. B. Taylor, *Annu. Rev. Astron. Astrophys.* **40**, 319 (2002).
6. T. E. Clarke, P. P. Kronberg, H. Böhringer, *Astrophys. J.* **547**, L111 (2001).
7. J. Bagchi, V. Pislar, G. B. Lima Neto, *Mon. Not. R. Astron. Soc.* **296**, L23 (1998).
8. G. Giovannini, L. Feretti, *J. Korean Astron. Soc.* **37**, 323 (2004).
9. T. A. Enßlin, P. L. Biermann, U. Klein, S. Kohle, *Astron. Astrophys.* **332**, 395 (1998).
10. We have used a Hubble constant of $H_0 = h$ 70 km/(s Mpc), where h is a dimensionless scaling parameter. We adopt a "concordance" model of big bang cosmology: the Λ-CDM (lambda cold dark matter) cosmological model defined by matter (dark matter + baryonic matter) density $\Omega_M = 0.3$, and dark energy (denoted by cosmological constant Λ) density $\Omega_\Lambda = 0.7$. Thus, $\Omega_M + \Omega_\Lambda = 1$. The Einstein–de Sitter cosmology has no dark energy, i.e. $\Omega_\Lambda = 0$, but $\Omega_M = 1$. All the dimensionless densities (Ω_M, Ω_Λ) are defined relative to the critical density for closure of the universe, $\rho_{crit} = (3\, H_0^2)/(8\pi G)$, where G is the gravitational constant. The Λ-CDM cosmology results in 0.904 kpc per arc sec plate-scale for redshift $z = 0.046$.
11. R. J. Smith *et al.*, *Astron. J.* **128**, 1158 (2004).
12. Materials and methods are available as supporting material on *Science* Online.
13. E. Escalera *et al.*, *Astrophys. J.* **423**, 539 (1994).
14. P. Flin, J. Krywult, *Astron. Astrophys.* **450**, 9 (2006).
15. H. Rottgering, M. H. Wieringa, R. W. Hunstead, R. D. Eckers, *Mon. Not. R. Astron. Soc.* **290**, 577 (1997).
16. M. Bruggen *et al.*, *Astrophys. J.* **631**, L21 (2005).
17. A. P. Schoenmakers, A. G. de Bruyn, H. J. A. Rottgering, H. van der Laan, *Astron. Astrophys.* **374**, 861 (2001).
18. F. Miniati, *Mon. Not. R. Astron. Soc.* **342**, 1009 (2003).
19. U. Keshet, E. Waxman, A. Loeb, V. Springel, L. Hernquist, *Astrophys. J.* **585**, 128 (2003).
20. P. M. Ricker, C. L. Sarazin, *Astrophys. J.* **561**, 621 (2001).
21. M. Hoeft, M. Bruggen, G. Yepes, *Mon. Not. R. Astron. Soc.* **347**, 389 (2004).
22. K. Roettiger, J. O. Burns, J. M. Stone, *Astrophys. J.* **518**, 603 (1999).
23. T. Kitayama, S. Yasushi, *Astrophys. J.* **469**, 480 (1996).
24. M. Girardi, G. Giuliano, F. Mardirossian, M. Marino, W. Boschin, *Astrophys. J.* **505**, 74 (1998).
25. S. Gabici, P. Blasi, *Astrophys. J.* **583**, 695 (2003).
26. H. Kang, T. W. Jones, *Astrophys. J.* **620**, 44 (2005).
27. The minimum-energy magnetic field B_{min} is obtained by minimizing the total nonthermal energy density (u_{min}), which consists of the magnetic-field energy and the energy contained in all relativistic particles. Specifically, $B_{min} \propto (1 + k)^{2/7} (\phi V)^{-2/7} L_{syn}^{2/7}$. In this expression, the model parameters are as follows: k, the ratio of energy in heavy particles to electrons; φ, the fraction of the source volume filled by magnetic fields; and V, the total source volume. The total synchrotron radio luminosity, L_{syn}, can be obtained from observations by integrating the radio spectrum between lowest and highest frequencies (we used 10 MHz to 100 GHz). Parameters K, φ, and V are usually unknown and need to be guessed reasonably. For our estimate of B_{min}, we used the values within range: K = 1 to 100, φ = 0.1 to 1, line-of-sight depth = 270 kpc (for calculating V), and spectral index α = 0.5 (for obtaining L_{syn}). For further details, see (*31*).
28. R. Kulsrud, R. Cen, J. P. Ostriker, D. Ryu, *Astrophys. J.* **480**, 481 (1997).
29. P. P. Kronberg, H. Lesch, U. Hopp, *Astrophys. J.* **511**, 56 (1999).
30. F. Miniati, D. Ryu, H. Kang, T. W. Jones, *Astrophys. J.* **559**, 59 (2001).
31. F. Govoni, L. Feretti, *Int. J. Mod. Phys.* (*D*) **13**, 1549 (2004).

32. F.D. acknowledges support from Centre National d'Etudes Spatiales and Programme National Cosmologie, CNRS/Institut National des Sciences de l'Univers. G.B.L.N. acknowledges support from Conselho Nacional de Desenvolvimento Científico e Technológico, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/Cofecub Brazilian-French collaboration, and Fundação de Amparo à Pesquisa de São Paulo through the Thematic Project 01/07342-7. S.P. acknowledges the Deutsche Forschungsgemeinschaft (DFG), Research Training Groups (RTG) 1147 for financial support and specially thanks The Inter-University Centre for Astronomy and Astrophysics for support. The National Radio Astronomy Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. The x-ray data are based on observations obtained with XMM-Newton, a European Space Agency (ESA) science mission with instruments and contributions directly funded by ESA Member States and NASA.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/791/DC1
SOM Text
Figs. S1 to S3
References and Notes

12 June 2006; accepted 6 October 2006
10.1126/science.1131189

Imaging the Sublimation Dynamics of Colloidal Crystallites

J. R. Savage,[1] D. W. Blair,[1] A. J. Levine,[2] R. A. Guyer,[1] A. D. Dinsmore[1]

We studied the kinetics of sublimating crystals with single-particle resolution by experiments with colloidal spheres and by computer simulations. A short-range attraction between spheres led to crystallites one to three layers thick. The spheres were tracked with optical microscopy while the attraction was reduced and the crystals sublimated. Large crystallites sublimated by escape of particles from the perimeter. The rate of shrinkage was greatly enhanced, however, when the size decreased to less than 20 to 50 particles, depending on the location in the phase diagram. At this size, the crystallites transformed into a dense amorphous structure, which rapidly vaporized. The enhancement of kinetics by metastable or unstable phases may play a major role in the melting, freezing, and annealing of crystals.

The process by which crystals melt into a liquid or sublimate into a gas are of considerable fundamental and practical importance. Even in model systems whose equilibrium configurations are known, the kinetics of phase transitions have proved difficult to understand and predict (*1*–*4*). In the case of nucleation of a crystalline phase from a fluid, the kinetics are slowed by the energy cost of forming an interface between the new phase and the old one, as described by classical nucleation theory. Complicating the process, thermodynamically metastable or unstable phases enhance the nucleation rate by lowering the free-energy barrier that must be overcome to reach equilibrium (*5*–*10*). In studies of the reverse process—melting or sublimation—two different mechanisms have been identified. Three-dimensional crystallites that have stabilized surfaces melt from within through mechanical or Born melting (*11*, *12*) once the shear modulus is reduced to nearly zero (*11*–*14*). In this case, melting might be viewed as nucleation and growth of the fluid phase within the solid (*11*, *12*, *14*). In two dimensions, large crystallites melt by a two-step process owing to the intervention of the hexatic phase, which appears in equilibrium (*3*, *15*–*19*). In many practical cases, however, crystallites are of finite size. Melting or sublimation then take place at the surface (or perimeter) while the interior retains its crystalline order (*2*, *3*). Thus, in experiments with real systems of melting or sublimating crystallites, the melting process is dominated by the surface (*20*). As crystallites shrink in size, one might expect the surface to play a continually greater role in the kinetics. Owing to the difficulty of studying small crystallites with single-particle resolution, however, this has remained an open question. Experiments with colloidal crystals at very low ionic strength have shown unexpectedly slow melting, but this has been attributed to a long-range attraction (*21*). We found that crystallites sublimate at a steady rate from their perimeter until they reach a characteristic size, after which they very rapidly vaporize.

[1]Department of Physics, University of Massachusetts Amherst, Hasbrouck Lab 411, 666 North Pleasant Street, Amherst, MA 01003, USA. [2]Department of Chemistry and Biochemistry and California Nanosystems Institute, University of California, Los Angeles, CA 90095, USA.

Images of a typical colloidal crystallite while it sublimates into a low-density gas phase are shown in Fig. 1. The video images show colloidal spheres in water containing surfactant micelles (not visible). The micelles induce an attractive potential between spheres by the depletion, or excluded-volume, mechanism (*22*). The range of the depletion potential is determined by the diameter of the micelles, which is approximately 0.3 to 1.4% of the sphere diameter. The key to our experiment is that the size and concentration of the micelles—and hence the strength of attraction—are strong functions of temperature, *T*. Using the nonionic surfactant hexaethylene glycol monododecyl ether ($C_{12}E_6$), we obtained crystallites of the colloidal spheres coexisting with a gas phase at 28°C. These crystallites form at a flat glass surface owing to depletion attraction to the surface (*23*); most crystallites contain only a single layer of spheres. Cooling the sample to 27°C weakens the depletion attraction (*22*) and the crystallites sublimate (Fig. 1). During this transition, almost all of the spheres remain on the surface and diffuse in two dimensions. The images clearly indicate a two-stage process, in which large crystallites shrink by departure of particles from the perimeter monolayer at a rate limited by breaking of bonds (Fig. 1, A to C, time Δt = 100 s between images), whereas smaller clusters melt much more rapidly (Fig. 1, D to F, Δt = 15 s).

To quantify the kinetics, Fig. 2 shows the cluster size $N(t)$ for several crystallites. Figure 2A shows data for three colloidal crystallites formed in the presence of $C_{12}E_6$ at 28°C and then cooled to 27°C, similar to that in Fig. 1. Here, N is defined as the number of particles that are connected in a contiguous cluster by depletion bonds, which are defined by their separation (*22*). This definition of cluster is independent of the degree of crystalline order. For large clusters or early time, the rate of sublimation varies among crystallites, perhaps owing to variations in their shape. When N falls below about 30, however, the clusters are more isotropic in shape and in each case the rate markedly increases. We discerned no dependence of the final shrinkage rate on the cluster's history.

The results shown in Figs. 1 and 2A are robust; we observed similar results in other experiments and in computer simulations. Figure 2B shows data for four colloidal crystals formed in the presence of micelles of an anionic surfactant, sodium dodecyl sulfate (SDS), at room temperature. Two of these melted at 30° and two melted at 65°C; unlike the $C_{12}E_6$, the SDS micelles shrink with increasing T (*22*). Figure S1 shows $N(t)$ for all seven crystallites that we studied at 65°C. The typical rate of crystallite shrinkage for $N > 30$ increases with T by an order of magnitude, yet all samples show the break in the slope at $N \sim 20$ to 30. Figure 2C shows two crystallites melted in a Brownian dynamics computer simulation (*22*) of a quasi–two-dimensional system. The attractive pair potential approximated the depletion potential with a range equal to 10% of the particle diameter. These data adhered to the same enhanced kinetics at small N (≈50 in simulation), which excludes hydrodynamic flow as an explanation of our experimental results. Experimental data for colloidal crystals of two to three layers in thickness (fig. S2) also showed enhanced kinetics at small size, showing that the cross-over was not a strictly two-dimensional effect. All of these melting curves clearly showed a tendency to sublimate at a steady rate until N reached a characteristic size. In crystallites below this size, the melting rate was greatly enhanced.

If the melting rate were limited by the rate of thermally breaking bonds at the perimeter, then $dN/dt \propto N^{1/2}$ and the melting rate should slow down rather than speed up. Alternatively, classical nucleation theory predicts a critical size for a growing

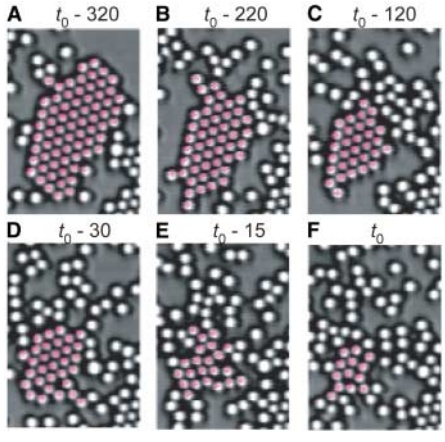


Fig. 1. Images of a crystallite of 0.7-μm-radius polystyrene spheres during sublimation. The purple dots indicate particles identified by the cluster algorithm. Images were obtained using an inverted optical microscope (63× magnification, 1.3 numerical aperture) and a monochrome video camera. Times (in seconds) are measured relative to t_0, when the crystallites were defined as having vanished. The images indicate a relatively slow rate initially [Δt = 100 s (**A** to **C**)] and an enhanced rate near the end [Δt = 15 s (**D** to **F**)]. The sample was heated from above and below with feedback-controlled resistive heaters. A sapphire plate was mounted to the top surface to suppress temperature gradients. The temperature had equilibrated and random fluctuations in temperature were approximately 0.1°C.

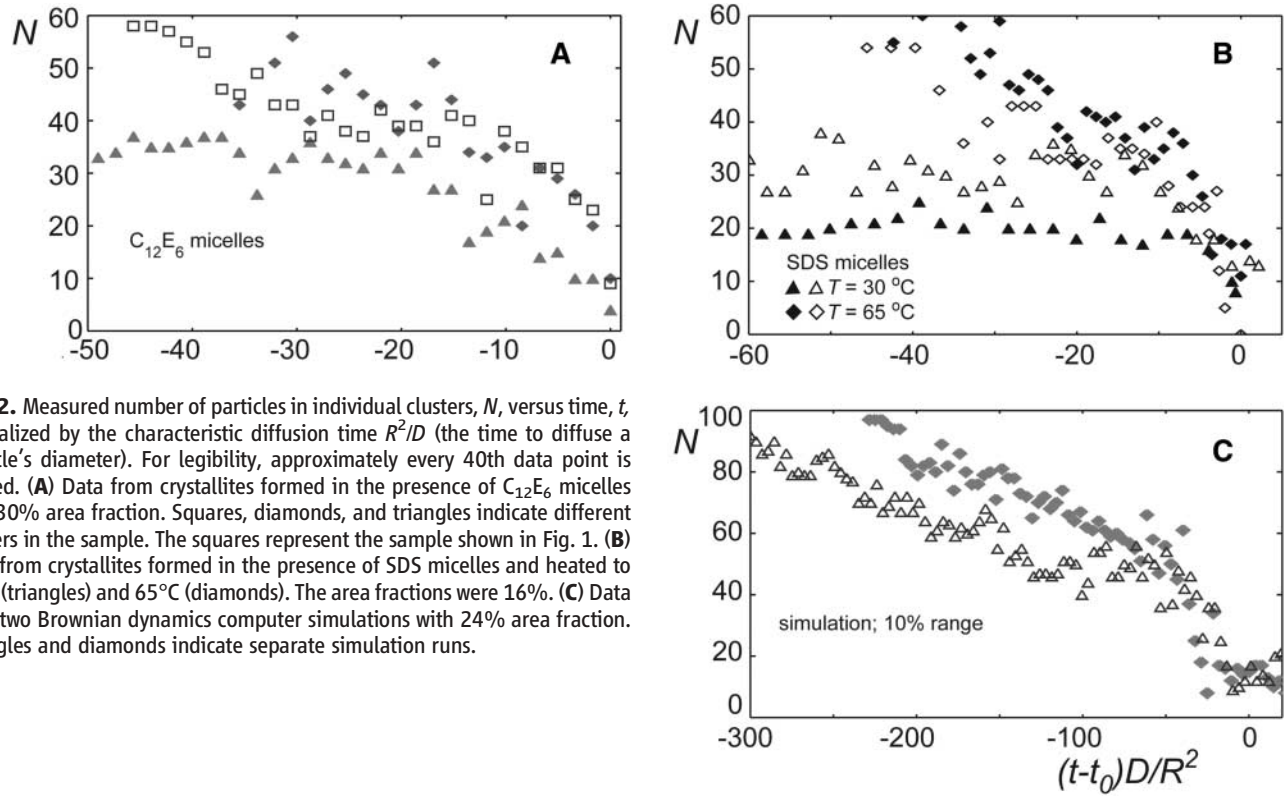


Fig. 2. Measured number of particles in individual clusters, *N*, versus time, *t*, normalized by the characteristic diffusion time R^2/D (the time to diffuse a particle's diameter). For legibility, approximately every 40th data point is plotted. (**A**) Data from crystallites formed in the presence of $C_{12}E_6$ micelles with 30% area fraction. Squares, diamonds, and triangles indicate different clusters in the sample. The squares represent the sample shown in Fig. 1. (**B**) Data from crystallites formed in the presence of SDS micelles and heated to 30°C (triangles) and 65°C (diamonds). The area fractions were 16%. (**C**) Data from two Brownian dynamics computer simulations with 24% area fraction. Triangles and diamonds indicate separate simulation runs.

crystallite, which is reminiscent of the characteristic size reported here. However, in nucleation theory this size arises from a competition between the lower chemical potential of the crystal phase and its positive surface energy. During melting, the chemical potential is larger in the crystal, both terms are positive, and no critical size emerges. Instead, our analysis points to a third mechanism. We show that the cross-over in kinetics arises because of a two-step process: (i) The crystallite forms a dense amorphous phase, which is thermodynamically unstable, and (ii) it rapidly evaporates into the gas phase.

To distinguish crystalline from amorphous clusters, we measured the sixfold symmetry of bonds within the clusters using the bond-orientation order parameter, ψ_6 (*15*). For each particle (labeled *j*) that has two or more bonds, $\psi_6(j) \equiv (1/Z)\Sigma_k \exp[6i\theta_{jk}]$, where *Z* is the number of bonds with this particle, the sum is over all bonded neighbors *k*, and θ_{jk} is the angle between the *j*–*k* bond and the *x* axis. Computing the modulus squared, $|\psi_6|^2$ eliminates the arbitrary choice of axis. $|\psi_6|^2 = 1$ in a perfect hexagonal lattice, independent of the number of bonds. We found that the downturn in $N(t)$ is simultaneous with a drop in $|\psi_6|^2$ from between 0.8 and 1 to about 0.2.

To focus on the effect of cluster size *N* rather than time, we plotted $|\psi_6|^2_{av}$ versus *N*, where the subscript "av" refers to an average over all clusters of size *N* (Fig. 3A). For colloidal crystallites formed in the presence of $C_{12}E_6$ micelles and melting at 27°C (squares), $|\psi_6|^2_{av}$ is notably reduced when *N* is smaller than a cross-over value, $N_x \approx 30$. A similar result was found in the presence of SDS micelles at 65°C (diamonds) and at 30°C (triangles), although in the latter case N_x is closer to 20. In the computer simulation, with a range of attraction equal to 10% of the particle diameter, N_x 50 (Fig. 3D).

Even more revealing of the loss of orientational order is the product of each particle's ψ_6 value with that of its neighbors, which we call C_6 (*21*) (Fig. 3B). In a dense fluid, particles may have a high degree of sixfold bond order; unlike in a crystal, however, the directions of the bonds vary from one region to another. Hence, the thermal average of the product of $\psi_6(i) \times \psi_6(j)^*$ vanishes when particles *i* and *j* are far apart. Indeed, long-range bond orientational (ψ_6) order is a hallmark of two-dimensional crystals, distinguishing them from hexatic and fluid phases, which exhibit quasi–long-range (algebraically decaying) and short-range order, respectively (*15*). As expected, particles with $Z < 6$ (which are near the perimeter) have lower values of C_6 than do their neighbors with $Z = 6$ (Fig. 3B). As crystallites sublimate, an increasing fraction, *f*, of the particles have $Z < 6$ (Fig. 3C). Notably, the crystallites do not lose order simply because *f* increases, as might be expected if melting took place at a wetting layer. Instead, the particles with $Z < 6$ (perimeter particles) themselves lose orientational order when $N < N_x$. Particles with $Z = 6$ and one layer in become very rare, but these also show reduced order (though the effect is small because packing six neighbors at short range requires nearly 60° bond angles). Hence, small clusters have lower bond-orientational order throughout their area and should be characterized as amorphous.

These results point to the following mechanism for melting. Clusters larger than the cross-over size N_x sublimate directly into the gas phase from the perimeter region. Here, particles at the perimeter typically have lower values of $|\psi_6|^2$ and C_6. In the simulations, we found that mean-square fluctuations in the bond lengths are larger at the surface, as in earlier studies of large crystallites (*4*, *24*). Particles that lie one or more layer from the perimeter region tend to remain within the cluster and have C_6 values close to 1 (Fig. 3B). On the other hand, as *N* falls below N_x, a cluster rapidly forms a dense amorphous (liquid) phase. Here, $|\psi_6|^2_{av}$ and C_6 are small throughout the cluster area, which is the signature of a liquid. The liquid phase is thermodynamically unstable, and thus rapidly evaporates into the surrounding gas phase. Although this liquid phase is not found in equilibrium, it has a profound effect on the transition kinetics.

Systems with short-range attraction can also exhibit a second solid phase, which has the same symmetry as the other crystal but lower density (*25*). This expanded solid is expected to become metastable in the presence of the 3.5% size polydispersity used here (*26*). Even when the expanded solid is metastable, however, simulations showed that crystallites nucleate with the density of the (metastable) expanded phase, then contract to the stable-solid density as they grow (*27*). Correspondingly, we found that our sublimating crystallites dilate slightly as *N* decreases toward N_x. Our data, however, do not directly reveal two solid phases; we therefore focus on the more apparent amorphous phase.

To verify the role of a liquid phase in the sublimation kinetics, we performed simulations in the region of the phase diagram where gas-liquid coexistence is found in equilibrium. In accord with

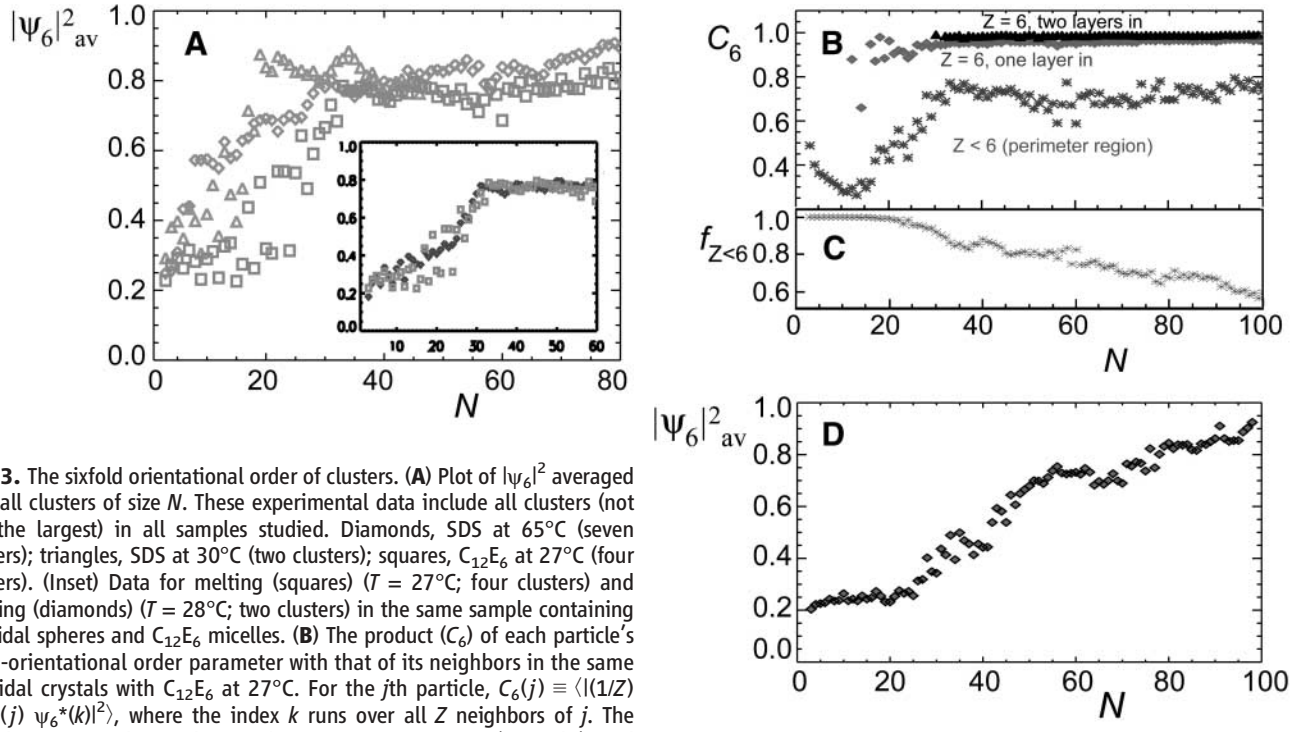


Fig. 3. The sixfold orientational order of clusters. (**A**) Plot of $|\psi_6|^2$ averaged over all clusters of size N. These experimental data include all clusters (not just the largest) in all samples studied. Diamonds, SDS at 65°C (seven clusters); triangles, SDS at 30°C (two clusters); squares, $C_{12}E_6$ at 27°C (four clusters). (Inset) Data for melting (squares) (T = 27°C; four clusters) and freezing (diamonds) (T = 28°C; two clusters) in the same sample containing colloidal spheres and $C_{12}E_6$ micelles. (**B**) The product (C_6) of each particle's bond-orientational order parameter with that of its neighbors in the same colloidal crystals with $C_{12}E_6$ at 27°C. For the jth particle, $C_6(j) \equiv \langle |(1/Z) \Sigma_k \psi_6(j)\ \psi_6^*(k)|^2 \rangle$, where the index k runs over all Z neighbors of j. The particles are sorted according to their position: perimeter (asterisks), and first (diamonds) and second (triangles) layers in from the perimeter. The perimeter particles are defined as those with fewer than six bonds; the first-layer particles have six bonds and are contiguous with the perimeter. The second-layer particles are contiguous with the first layer. (**C**) The fraction, f, of particles that have $Z < 6$. (**D**) $|\psi_6|^2$ averaged over all clusters of size N for a computer simulation (averaged over six runs).

earlier investigations in two and three dimensions (*4*, *6*, *28*–*30*), our quasi–two-dimensional simulations reveal only one amorphous phase (gas) in equilibrium when the attraction is of short range. When the range of attraction exceeds approximately 60% of the diameter, however, we found coexistence of gas and liquid phases in equilibrium (*22*). Whereas short-range attraction leads to a transient liquid phase, a long-range attraction leads to a stable liquid. For example, when the range of attraction is 80% of the sphere diameter, a crystallite of $N = 100$ rapidly forms a dense liquid droplet, which shrinks to N 85 and then persists for at least 1000 R^2/D, where R is the particle radius and D is the diffusion constant. Earlier reports of the melting of ligand-stabilized CdS nanoparticles in vacuum have also found that the crystalline nanoparticles melt to form stable liquid droplets upon heating (*31*). Apparently, as our clusters shrink in size, the free energy of the solid phase approaches and then exceeds that of the liquid [which can also be inferred from the evidence that the melting temperature decreases for small crystallites (*31*–*33*)]. At this point, forming a liquid from the crystal reduces the energy. Because relatively few bonds need to be broken in order to disrupt the long-range order of the crystallite, this process is plausibly fast. Once the liquid forms, the static shear modulus vanishes; if this liquid is thermodynamically unstable, thermal fluctuations cause rapid evaporation.

Notably, we found a similar dependence of $|\psi_6|^2$ and C_6 on N during freezing (inset of Fig. 3A). For colloidal crystals in the presence of $C_{12}E_6$, the sample was subjected to repeated freezing and melting by repeated switching of T from 28° to 27° C (waiting sufficient time between cycles to ensure complete melting and looking in different regions of the sample). During the repeated cycles, crystallites showed statistically indistinguishable $|\psi_6|^2_{av}$ versus N. Hence, we conclude that the liquid phase, which is unstable or metastable in this case, is intimately involved in nucleation as well as melting. This conclusion agrees with earlier computer simulations (*6*), density functional theory (*8*, *10*), and experiments (*9*) in three dimensions.

The results described here were found in crystallites of one to three layers thick, but are likely to be generally applicable to crystallites of arbitrary thickness formed from centrosymmetric, attractive potentials. Short-range attractions might prevail in crystallization and melting of macromolecules such as proteins in membranes or in three dimensional solutions; in these cases, the liquid phase is metastable or unstable and small crystallites would rapidly vanish. Crystallites of atoms with longer-range (Lennard-Jones–like) attraction are common in vacuum-phase film deposition and as nanoparticles on substrates or in solution; in these cases, the liquid phase is stable and small crystallites can form tiny liquid droplets that persist, as indicated by our simulations. In clusters composed of nonspherical particles, other metastable phases with body-centered cubic (*34*, *35*) or liquid crystalline symmetry (*36*) might determine the kinetics. Hence, the observed two-stage sublimation process should be generally relevant to the melting, freezing, and annealing kinetics of small clusters with short-range attraction.

References and Notes

1. A. Laaksonen, V. Talanquer, D. W. Oxtoby, *Annu. Rev. Phys. Chem.* **46**, 489 (1995).
2. R. W. Cahn, *Nature* **413**, 582 (2001).
3. J. G. Dash, *Rev. Mod. Phys.* **71**, 1737 (1999).
4. V. J. Anderson, H. N. W. Lekkerkerker, *Nature* **416**, 811 (2002).
5. J. W. Cahn, *J. Am. Ceram. Soc.* **52**, 118 (1969).
6. P. R. ten Wolde, D. Frenkel, *Science* **277**, 1975 (1997).
7. R. M. L. Evans, W. C. K. Poon, M. E. Cates, *Europhys. Lett.* **38**, 595 (1997).
8. V. Talanquer, D. W. Oxtoby, *J. Chem. Phys.* **109**, 223 (1998).
9. L. F. Filobelo, O. Galkin, P. G. Vekilov, *J. Chem. Phys.* **123**, 014904 (2005).
10. J. F. Lutsko, G. Nicolis, *Phys. Rev. Lett.* **96**, 046102 (2006).
11. K. Lu, Y. Li, *Phys. Rev. Lett.* **80**, 4474 (1998).
12. Z. H. Jin, P. Gumbsch, K. Lu, E. Ma, *Phys. Rev. Lett.* **87**, 055703 (2001).
13. M. Born, *J. Chem. Phys.* **7**, 591 (1939).
14. Y. J. Zhou, X. F. Jin, *Phys. Rev. B* **71**, 224113 (2005).
15. B. I. Halperin, D. R. Nelson, *Phys. Rev. Lett.* **41**, 121 (1978).
16. C. A. Murray, D. H. van Winkle, *Phys. Rev. Lett.* **58**, 1200 (1987).
17. A. H. Marcus, S. A. Rice, *Phys. Rev. Lett.* **77**, 2577 (1996).
18. R. A. Quinn, J. Goree, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **64**, 051404 (2001).
19. H. H. von Grunberg, P. Keim, K. Zahn, G. Maret, *Phys. Rev. Lett.* **93**, 255703 (2004).
20. R. W. Cahn, *Nature* **323**, 668 (1986).
21. A. E. Larsen, D. G. Grier, *Nature* **385**, 230 (1997).
22. Materials and methods are available as supporting material on *Science* Online.
23. P. D. Kaplan, J. L. Rouke, A. G. Yodh, D. J. Pine, *Phys. Rev. Lett.* **72**, 582 (1994).
24. A. M. Alsayed, M. F. Islam, J. Zhang, P. J. Collings, A. G. Yodh, *Science* **309**, 1207 (2005).

25. P. Bolhuis, D. Frenkel, *Phys. Rev. Lett.* **72**, 2211 (1994).
26. P. G. Bolhuis, Ph.D. thesis, University of Utrecht, Netherlands (1996).
27. A. Cacciuto, S. Auer, D. Frenkel, *Phys. Rev. Lett.* **93**, 166105 (2004).
28. H. N. W. Lekkerkerker, W. C. K. Poon, P. N. Pusey, A. Stroobants, P. B. Warren, *Europhys. Lett.* **20**, 559 (1992).
29. S. M. Ilett, A. Orrock, W. C. K. Poon, P. N. Pusey, *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **51**, 1344 (1995).
30. J. F. Lutsko, G. Nicolis, *J. Chem. Phys.* **122**, 244907 (2005).
31. A. N. Goldstein, C. M. Echer, A. P. Alivisatos, *Science* **256**, 1425 (1992).
32. P. R. Couchman, C. L. Ryan, *Philos. Mag. A* **37**, 369 (1978).
33. M. Wautelet, *J. Phys. D* **24**, 343 (1991).
34. P. R. ten Wolde, M. J. Ruiz-Montero, D. Frenkel, *Phys. Rev. Lett.* **75**, 2714 (1995).
35. Y. C. Shen, D. W. Oxtoby, *Phys. Rev. Lett.* **77**, 3585 (1996).
36. A. M. Alsayed, Z. Dogic, A. G. Yodh, *Phys. Rev. Lett.* **93**, 057801 (2004).
37. We gratefully acknowledge J. Machta and D. Frenkel for helpful discussions. J.R.S. and A.D.D. acknowledge support from the Research Corporation and NSF grant DMR-0305395. D.W.B. acknowledges support from NSF grant DMR-0242402, and A.J.L. acknowledges support from the NASA Fluid Physics Program (02-OBPR-03-C).

Supporting Online Material
www.sciencemag.org/cgi/content/full/314/5800/795/DC1
Materials and Methods
Figs. S1 to S4
References

12 April 2006; accepted 29 August 2006
10.1126/science.1128649

Pattern Formation and Shape Changes in Self-Oscillating Polymer Gels

Victor V. Yashin and Anna C. Balazs*

We developed an efficient model for responsive gels that captures large-scale, two-dimensional (2D) deformations and chemical reactions within a swollen polymer network. The 2D calculations allowed us to probe not only volume changes but also changes in sample shape. By focusing on gels undergoing the oscillatory Belousov-Zhabotinsky reaction, we observed traveling waves of local swelling that form a rich variety of dynamic patterns and give rise to distinctive oscillations in the gel's shape. The observed patterns depend critically on the gel's dimensions. The approach provides a useful computational tool for probing the dynamics of chemomechanical processes and uncovering morphological transformations in responsive gels.

For a synthetic material to perform sustained mechanical work, it must undergo large-scale, periodic changes in volume or shape. Polymeric gels constitute optimal candidates for use as soft active materials, because controlled modulations of the surrounding solvent can lead to significant, rhythmic expansion and contraction of the gel (*1*, *2*). Consequently, such oscillating gels could be used as microactuators (*3*) for pulsatile drug delivery (*4*). In addition to their practical utility, oscillating gels provide an ideal medium for investigating nonlinear dynamical phenomena that can arise from a coupling of mechanical and chemical energy. For example, researchers have isolated scenarios where the initial swelling and deswelling of a chemoresponsive gel exerts feedback on a nonoscillatory chemical reaction and thereby drives the entire system into a regular, oscillatory regime (*4*–*8*). Theoretical models for oscillating gels (*2*, *7*–*9*) have yielded substantial insight into mechanisms that can produce periodic pulsations. Prior calculations (*2*, *7*–*9*), however, were carried out in one dimension (e.g., the sample was assumed to be spherically symmetric), and thus only volumetric changes of the sample could be probed. To capture shape changes, models must describe the gel deformation in at least two dimensions. By encompassing additional degrees of freedom, 2D (or 3D) models can enhance our fundamental understanding of the interplay between the finite deformations of a responsive medium and the nonlinear chemical dynamics and open up the possibility of uncovering new morphological transitions (*7*).

We developed an approach for simulating chemoresponsive gels that can exhibit not only large variations in volume but also alterations in shape. Through this approach, we modeled oscillating gels undergoing the Belousov-Zhabotinsky (BZ) reaction (*10*–*12*) and showed that the ensuing 2D pattern formation depends on the dimensions of the sample. BZ gels are unique because the polymer network can expand and contract periodically without external stimuli (*11*, *12*). This autonomous, self-oscillatory behavior is due to a ruthenium catalyst, which is covalently bonded to the polymers (*11*, *12*). The BZ reaction generates a periodic oxidation and reduction of the anchored metal ion, and these chemical oscillations induce the rhythmic swelling-deswelling in the gel. By probing morphological changes in BZ gels, we can establish design criteria for creating autonomous small-scale devices, which perform sustained work until the reagents in the host solution are consumed and can be simply refueled by replenishing these solutes.

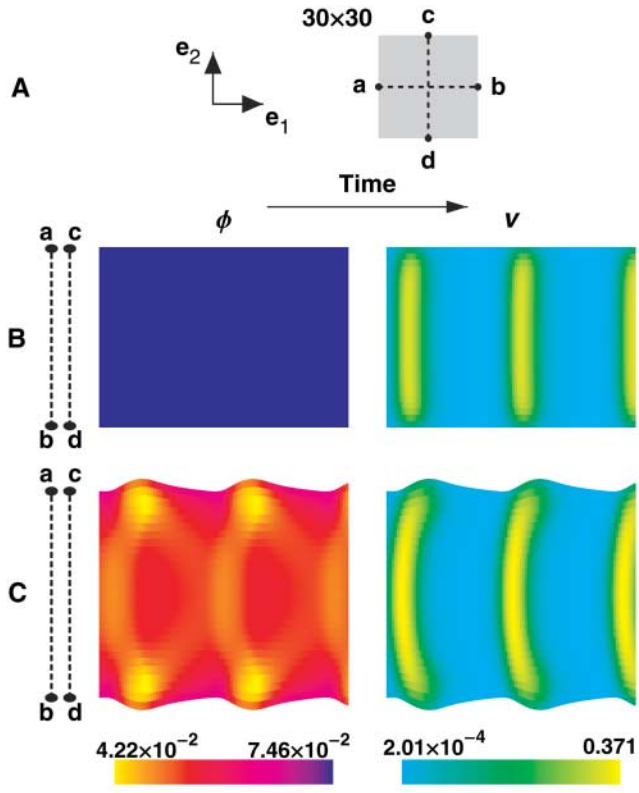


Fig. 1. Temporal evolution of ϕ and v in a 30-by-30 square gel. (**A**) Location of horizontal (a-b) and vertical (c-d) cross sections through the sample. (**B**) Nonresponsive gel ($\chi^* = 0$). (**C**) Responsive gel ($\chi^* = 0.105$). Rhythmic variation of gel size is clearly seen. Simulations were performed with (*9*): $\phi_0 = 0.139$, $c_0 v_0 = 1.3 \times 10^{-3}$, $\chi(\phi) = 0.338 + 0.518\,\phi$, $\lambda_\perp = 1.1$, $\varepsilon = 0.354$, $f = 0.7$, and $q = 9.52 \times 10^{-4}$.

Chemical Engineering Department, University of Pittsburgh, Pittsburgh, PA 15261, USA.

*To whom correspondence should be addressed. E-mail: balazs1@engr.pitt.edu

Our approach was inspired by the lattice spring model (LSM), where a material is represented by a network of interconnected Hookean springs (*13*); however, we made substantial modifications to create a gel lattice spring model, or gLSM, which captures the gel's large-scale deformations, interaction with the surrounding solvent, and response to chemical reactions. We started with a 2D lattice, which corresponds to a thin film or a 3D sample where the height of the material remains constant. In the initial undeformed system, this layer height is H_0, and the length of each element in the square lattice is Δ (fig. S1A). The spatial position of every node in this lattice is uniquely characterized by the vector $\mathbf{m} = k\ \mathbf{e}_1 + l\ \mathbf{e}_2$ or, equivalently, by the pair of integers (k,l). The elements are labeled by the pair $\mathbf{m} = (k,l)$ corresponding to the node in the lower left corner of the element. Upon deformation, the vectors $\mathbf{m}$ merely label the nodes and the elements, whereas the vectors $\boldsymbol{x}_\mathbf{m}$ give the actual positions of the nodes in the laboratory coordinate system (fig. S1A). The total energy of the deformed system, E, can be expressed as a function of the coordinates of the nodes, or $E = U(\{\boldsymbol{x}_\mathbf{m}\})$, where $U(\{\boldsymbol{x}_\mathbf{m}\}) = H_0\Delta^2\sum_\mathbf{m} u_\mathbf{m}$ is the sum of the energies of the deformed rectangular elements.

To capture the nonlinear elasticity of the gels, we must express the energy density, u, in terms of the invariants of the strain tensor, I_i (*14*). For purely 2D deformation, the invariants of interest are (*14*, *15*) $I_1 = \boldsymbol{d}_1^2 + \boldsymbol{d}_2^2 + \lambda_\perp^2$ and $I_3 = \lambda_\perp^2[\mathbf{e}_3 \cdot (\boldsymbol{d}_1 \times \boldsymbol{d}_2)]^2$, where $\boldsymbol{d}_i = \partial\boldsymbol{x}/\partial X_i$ for i values of 1 and 2, and $\lambda_\perp$ and $\mathbf{e}_3$ are the degree of swelling and the unit vector in the direction perpendicular to the deformation, respectively. The energy density $u = u(I_1, I_3)$ is a sum of two distinct contributions (*8*). The first, u_{el}, describes the rubber elasticity of the cross-linked polymer chains and can be expressed as (*16*)

$$\frac{u_{el}}{T} = \frac{c_0}{2}(I_1 - 3 - \ln I_3^{1/2}) \qquad (1)$$

where T is temperature (measured in energy units) and c_0 is the cross-link density, that is, the number density of elastic strands in the undeformed polymer network. In the undeformed state, u_{el} equals 0 because $I_1 = 3$ and $I_3 = 1$. The second contribution to the energy density, u_{FH}, describes the interaction between the polymer and solvent units:

$$u_{FH}/T = v_0^{-1} I_3^{1/2} f_{FH}(\phi_0 I_3^{-1/2}) \qquad (2)$$

where v_0 is the volume of a monomer within the chains, f_{FH} is the Flory-Huggins free energy density (*17*), and $f_{FH}(\phi) = (1-\phi)\ln(1-\phi) + \chi_{FH}\phi(1-\phi)$, with χ_{FH} being the polymer-solvent interaction parameter. The local volume fraction of polymer in the gel, ϕ, depends on the volume fraction of polymer in the undeformed state, ϕ_0, and the local deformation as $\phi = \phi_0 I_3^{-1/2}$.

The force acting on a node $\mathbf{m}$ can now be obtained as $\boldsymbol{F}_\mathbf{m} = -\partial U/\partial \boldsymbol{x}_\mathbf{m}$, which yields

$$\begin{aligned}\boldsymbol{F}_{(k,l)} = {} & c_0 T \frac{H_0}{2} \sum_{\delta,\delta'=\pm 1} [\boldsymbol{x}_{(k+\delta,l+\delta')} - \boldsymbol{x}_{(k,l)}] + \\ & \lambda_\perp T \frac{H_0}{2} \Big\{ \mathbf{e}_3 \times [\boldsymbol{x}_{(k,l+1)} - \boldsymbol{x}_{(k+1,l)}] P_{(k,l)} + \\ & \mathbf{e}_3 \times [\boldsymbol{x}_{(k-1,l)} - \boldsymbol{x}_{(k,l-1)}] P_{(k-1,l)} + \\ & \mathbf{e}_3 \times [\boldsymbol{x}_{(k,l-1)} - \boldsymbol{x}_{(k-1,l)}] P_{(k-1,l-1)} + \\ & \mathbf{e}_3 \times [\boldsymbol{x}_{(k+1,l)} - \boldsymbol{x}_{(k,l-1)}] P_{(k,l-1)} \Big\} \end{aligned} \qquad (3)$$

Here, $P_\mathbf{m}$ is the pressure in element $\mathbf{m}$, with $P_\mathbf{m} = \pi_\mathbf{m} + c_\mathbf{m} T/2$. In the latter equation, $\pi_\mathbf{m}$ is the Flory-Huggins osmotic pressure (*17*) in element $\mathbf{m}$:

$$\pi_\mathbf{m} = -v_0^{-1} T [\phi_\mathbf{m} + \ln(1-\phi_\mathbf{m}) + \chi(\phi_\mathbf{m})\phi_\mathbf{m}^2] \qquad (4)$$

where $\phi_\mathbf{m} = \phi_0 I_3^{-1/2}(\mathbf{m})$ is the volume fraction of polymer in element $\mathbf{m}$. Similarly, $c_\mathbf{m}$ is the local value of the cross-link density and is equal to $c_\mathbf{m} = c_0 I_3^{-1/2}(\mathbf{m})$. Because of the high friction between the polymer and solvent, inertial effects can be neglected, and the node velocity is proportional to the force acting on that site (*18*) so that

$$dx_\mathbf{m}/dt = \Lambda_\mathbf{m} F_\mathbf{m} \qquad (5)$$

The mobility $\Lambda_\mathbf{m} \propto H_0^{-1}\Delta^{-2}\zeta^{-1}(\phi_\mathbf{m})$ is inversely proportional to the polymer-solvent friction coefficient, ζ, where $\zeta(\phi) \propto \phi^{3/2}$ for a good solvent (*18*).

Equation 3 allows us to interpret the internal gel forces as a sum of two contributions: (i) the spring-like forces acting between nodes and (ii) the forces that originate from the pressure between the adjacent elements. A graphical representation of the equation for the force acting on node $\mathbf{m} = (k, l)$ is shown in fig. S1B. Note that one can readily introduce physical and chemical heterogeneities at different locations in the sample and thereby analyze how the local variations affect global properties.

This gLSM can be readily generalized to incorporate the BZ reaction. In addition to the local volume fraction of polymer, $\phi_\mathbf{m}$, and the cross-link density, $c_\mathbf{m}$, each element is now characterized by the dimensionless concentrations of the following (*9*, *19*): reactant dissolved

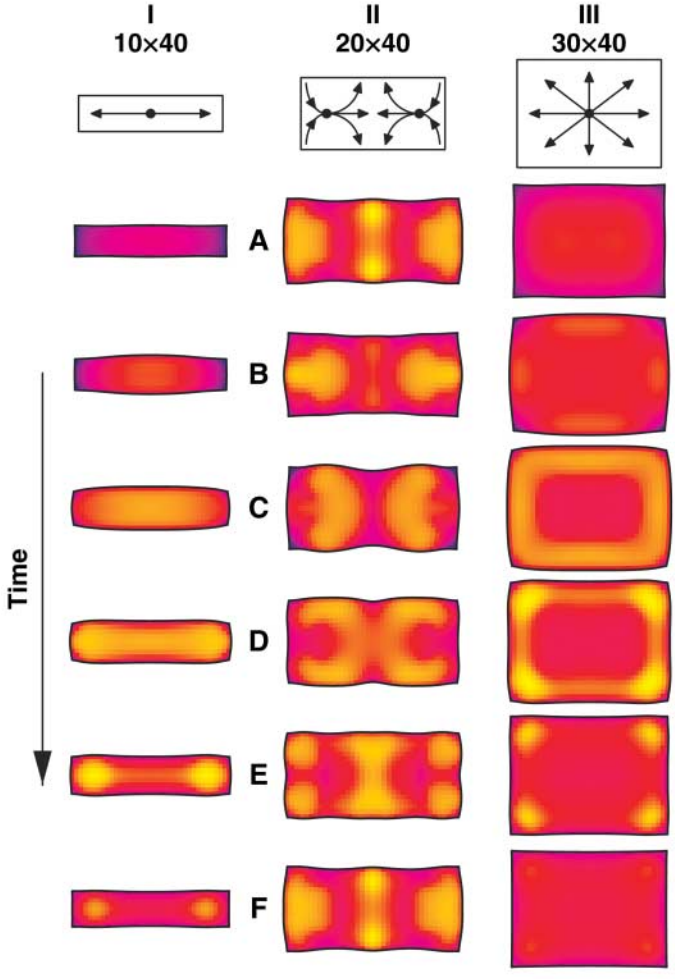


Fig. 2. Two-dimensional wave patterns and shape changes for gels of different dimensions. (**A** to **F**) Plots of ϕ during one period of oscillation for gel size (I) 10 by 40, (II) 20 by 40, and (III) 30 by 40; time step is $6T_0$. Initial degrees of swelling given by $\lambda_x = \lambda_y = \lambda_\perp$ and $u = v = 1 \times 10^{-3}$. All other parameters are as in Fig. 1.

in the solvent, $u_{\mathbf{m}}$, and oxidized metal-ion catalyst attached to the chains, $v_{\mathbf{m}}$. The gel dynamics are coupled to the BZ reaction via the pressure, $P_{\mathbf{m}}$, as

$$P_{\mathbf{m}} = \pi_{\mathbf{m}} + c_{\mathbf{m}}T/2 + v_0^{-1}T\chi^* v_{\mathbf{m}}\phi_{\mathbf{m}} \quad (6)$$

where χ^* is the coupling constant. When $\chi^* > 0$, the gel can respond to the chemical oscillations of the BZ reaction by undergoing spatial oscillations; when $\chi^* = 0$, the gel is nonresponsive to the chemical oscillations. For each element of the gLSM, the equations of chemical kinetics are formulated in terms of the mole fractions of the reagents, $\mathrm{u}_{\mathbf{m}} = u_{\mathbf{m}}\phi_{\mathbf{m}}^{-1}$ and $\mathrm{v}_{\mathbf{m}} = v_{\mathbf{m}}\phi_{\mathbf{m}}^{-1}$. Because the metal-ion catalyst is chemically linked to the polymer, the local chemical balance equation includes only the redox reaction:

$$d\mathrm{v}_{\mathbf{m}}/dt = \varepsilon\phi_{\mathbf{m}}^{-1}G(u_{\mathbf{m}}, v_{\mathbf{m}}, \phi_{\mathbf{m}}) \quad (7)$$

where ε is a constant. The other reactant undergoes both reaction and diffusion:

$$d\mathrm{u}_{\mathbf{m}}/dt = -\lambda_{\perp}\phi_0^{-1}\Delta^{-2}\sum_{i=1}^{4}\mathbf{e}_3\cdot[\mathbf{J}_i(\mathbf{m}) \times \xi_i(\mathbf{m})] + \phi_{\mathbf{m}}^{-1}F(u_{\mathbf{m}}, v_{\mathbf{m}}, \phi_{\mathbf{m}}) \quad (8)$$

The first term on the right-hand side of Eq. 8 describes the diffusive transport between the neighboring elements. $\mathbf{J}_i(\mathbf{m})$ is the flux of the dissolved reagent across edge i of the element $\mathbf{m}$, with $\xi_i(\mathbf{m})$ being the vector associated with this edge (fig. S1A). Both the transport of this reagent due to the polymer-solvent interdiffusion and the self-diffusion of the reagent in the solvent contribute to the flux $\mathbf{J}_i(\mathbf{m})$. G and F characterize the BZ reaction in the gel through the modified (*9*) Oregonator model (*19*):

$$F(u, v, \phi) = (1-\phi)^2 u - u^2 - fv(1-\phi) \times [u - q(1-\phi)^2][u + q(1-\phi)^2]^{-1} \quad (9)$$

$$G(u, v, \phi) = (1-\phi)^2 u - (1-\phi)v \quad (10)$$

where the constants f, ε, and q are the model parameters; a parameter accounting for the light sensitivity of the BZ reaction can be incorporated into Eq. 9 as outlined in (*20*). Equations 5, 7, and 8 are solved numerically to obtain $\phi_{\mathbf{m}}$, $u_{\mathbf{m}}$, and $v_{\mathbf{m}}$ as a function of time. We used the VODE code for solving large systems of stiff ordinary differential equations (*21*). The gel is freely swelling (i.e., no external forces act on the boundary nodes), and u equals 0 in the outer solution. Because the model is purely 2D, there are no transport processes in the third direction. The reaction rate functions F and G depend on ϕ, which depends on the local deformations; thus, the mechanical response of the system could affect the chemical kinetics of the system, as will be seen below.

The density plots in Fig. 1 show the temporal evolution of ϕ and the concentration of v for a particular portion of the gel. Specifically, we focused on the marked cross sections of a square sample (Fig. 1A) and plotted ϕ and v along these lines as a function of time. To highlight the difference between the responsive and nonresponsive gels, we first consider a system in which $\chi^* = 0$; that is, the polymer dynamics are not coupled to the oscillatory chemical reaction. The appearance of repeating, straight stripes in the v density plot (Fig. 1B) indicates that this variable exhibits periodic temporal oscillations (as dictated by the chosen values of f, ε, and q);

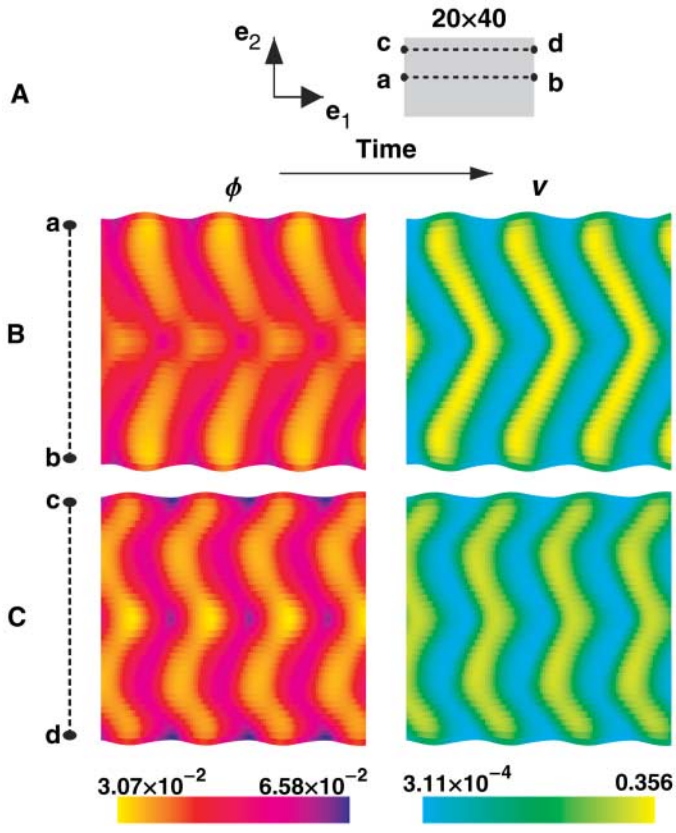


Fig. 3. Dynamic behavior along lateral cross sections of the 20-by-40 sample. (**A**) Location of cross sections. (**B**) Along a-b, the waves appear at the edges and propagate to the center of the sample. (**C**) Dynamics along c-d indicate the presence of two centers of wave generation. Period of time is $\delta t = 100\ T_0$.

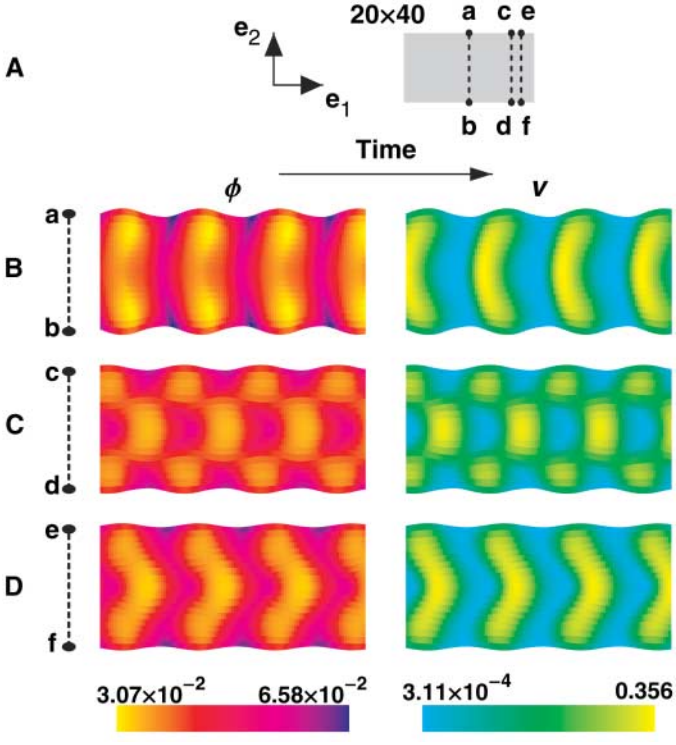


Fig. 4. Dynamic behavior along vertical cross sections for the 20-by-40 sample. (**A**) Location of cross sections. (**B**) The waves propagate from the center to the edges along a-b. (**C**) A standing wave is observed along c-d, which passes through one of the foci. (**D**) Along e-f, waves appear at the top and bottom and propagate inward.

however, the uniform coloring of the ϕ density plot reveals that the oscillatory changes in the catalyst have no effect on the chains.

When $\chi^* > 0$, we found dramatically different behavior. The coupling of the BZ reaction to the elasticity of the gel gives rise to traveling waves within the polymeric network, as seen by the waves of yellow in ϕ (Fig. 1C). The fact that these bent curves point to the left indicates that the waves originate from the center and propagate to the edges of the sample as time increases. In addition to this wave of local swelling, the BZ reaction modifies the shape of the responsive sample, as can be seen by the periodic variations in the sample's width. In this square example, similar oscillations appear in the vertical direction, causing the gel to undergo a symmetric, rhythmic expansion and contraction. The mechanical oscillations of the gel actually modify the BZ reaction, as can be seen by comparing *v* in Fig. 1, B and C. [It is worth noting that, for BZ reactions in solution, convective motion can lead to a feedback that alters the chemical reaction (*22*).]

We found that the dimensions of the sample have a notable effect on the dynamic patterns and shape changes that occur within the gel. Figure 2 shows snapshots in time for three examples: The sample width, *w*, is fixed at 40 lattice sites, whereas the height, *h*, is set to 10, 20, and 30 sites. The outer edges are marked in black to highlight the morphological transformations. For $h = 10$, the chemical waves lead to a swelling at the center of the slab; the swelling then propagates to the edges of the sample to yield a dog bone–shaped object. The entire process repeats, with the structure returning to the shape in column I, row A (IA). In this case, the waves emanate from the center and spread to the outer edges. This behavior also characterizes the patterns for the $h = 30$ sample, shown in column III. A swelling rapidly propagates to the edges, leaving an unswollen core in IIIF. The system then repeats the cycle, restarting with the configuration in IIIA.

The observed patterns are, however, much more intriguing when $h = 20$ (*23*). The waves appear to originate from two foci (Fig. 2, column II), merge in the center, and then decorate the edges, so that system resembles two back-to-back Cs. The outer edges of the Cs join to form the pattern in IIF, and the oscillations continue, returning to the pattern in IIA.

To further illustrate the distinct nature of the patterns for $h = 20$, we plotted the structural evolution of this system in both the lateral (Fig. 3) and vertical (Fig. 4) directions. Along the a-b cross section (Fig. 3A), the spatiotemporal pattern appears as repeating chevrons that point to the right (Fig. 3B). This pattern is indicative of waves that originate from the edges and propagate toward the center of the sample as time moves forward. In Fig. 3C, which shows the behavior along the c-d cross section, each bright line consists of two chevrons, corresponding to two waves that propagate away from the points of origin. When these waves reach the sample's center, they form a wave that propagates toward the edges in the vertical direction (Fig. 4B). If the vertical cross section passes through one of the two foci (Fig. 2 II), then the evolution exhibits a standing-wave pattern (Fig. 4C). Lastly, near the outer edges of the sample, the wave propagates inward again (Fig. 4D). Thus, the dynamic behavior for $h = 20$ is much more complex than those seen for $h = 10$ and $h = 30$, where the waves simply progress toward the edges, generating patterns similar to those in Fig. 1C. Although the aspect ratio h/w plays a role, further studies are needed to quantify how this value affects pattern formation.

To establish the relevant time and length scales, we noted that the unit of time in the dimensionless Eqs. 6 and 7 is T_0, the shortest time scale of the BZ reaction in solution (*19*, *24*). On the basis of rate constants in solution (*25*) and the typical concentrations in gels (*26*), we estimated $T_0 \sim 1$ s. The period of time in Fig. 2 is $\delta t = 100\ T_0$ or on the order of 1 to 2 min. The unit of length is $L_0 = (D_u T_0)^{1/2}$, where $D_u = 2 \times 10^{-5}$ cm^2 s^{-1} is the diffusion constant of the reactant *u* in solution (*27*), so the estimated length scale is $\Delta = L_0 \sim 40$ μm. These time and length scales are experimentally accessible, allowing the predictions to be readily tested.

By applying our gLSM technique to responsive gels undergoing the BZ reaction, we uncovered a rich variety of dynamic patterns and distinctive shape changes that depend on the dimensions of the sample. Additionally, within this system, chemical and mechanical behavior are highly coupled and mutually responsive. Overall, the studies further facilitate the design of devices that harness chemomechanical energy conversion to exhibit self-sustained rhythmic action in multiple directions.

References and Notes

1. M. Shibayama, T. Tanaka, *Adv. Polym. Sci.* **109**, 1 (1993).
2. J. A. Pojman, Q. Tran-Cong-Miyata, Eds., *Nonlinear Dynamics in Polymeric Systems* (ACS Symposium Series 869, American Chemical Society, Washington, DC, 2004).
3. J. R. Howse *et al.*, *Nano Lett.* **6**, 73 (2006).
4. A. P. Dhanarajan, G. P. Misra, R. A. Siegel, *J. Phys. Chem. A* **106**, 8835 (2002).
5. J.-C. Leroux, R. A. Siegel, *Chaos* **9**, 267 (1999).
6. V. Labrot, P. De Kepper, J. Boissonade, I. Szalai, F. Gauffre, *J. Phys. Chem. B* **109**, 21476 (2005).
7. J. Boissonade, *Phys. Rev. Lett.* **90**, 188302 (2003).
8. J. Boissonade, *Chaos* **15**, 023703 (2005).
9. V. V. Yashin, A. C. Balazs, *Macromolecules* **39**, 2024 (2006).
10. A. N. Zaikin, A. M. Zhabotinsky, *Nature* **225**, 535 (1970).
11. R. Yoshida, T. Takahashi, T. Yamaguchi, H. Ichijo, *J. Am. Chem. Soc.* **118**, 5134 (1996).
12. R. Yoshida, *Curr. Org. Chem.* **9**, 1617 (2005).
13. A. J. C. Ladd, J. H. Kinney, T. M. Breunig, *Phys. Rev. E* **55**, 3271 (1997).
14. R. J. Atkin, N. Fox, *An Introduction to the Theory of Elasticity* (Longman, New York, 1980).
15. A. Onuki, *Adv. Polym. Sci.* **109**, 63 (1993).
16. T. L. Hill, *An Introduction to Statistical Thermodynamics* (Addison-Wesley, Reading, MA, 1960).
17. P. G. DeGennes, *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, NY, 1979).
18. B. Barriere, L. Leibler, *J. Polym. Sci. Polym. Phys. Ed.* **41**, 166 (2003).
19. J. J. Tyson, P. C. Fife, *J. Chem. Phys.* **73**, 2224 (1980).
20. H.-J. Krug, L. Pohlmann, L. Kuhnert, *J. Phys. Chem.* **94**, 4862 (1990).
21. P. N. Brown, G. D. Byrne, A. C. Hindmarsh, *SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput.* **10**, 1038 (1989).
22. M. Diewald, K. Matthiessen, S. C. Muller, H. R. Brand, *Phys. Rev. Lett.* **77**, 4466 (1996).
23. The distinct patterns are observed only if the chemomechanical coupling is sufficiently strong; χ^* must lie above ~0.1 for the present choice of parameters.
24. J. J. Tyson, in *Oscillations and Traveling Waves in Chemical Systems*, R. J. Field, M. Burger, Eds. (Wiley, New York, 1985), p. 93.
25. T. Amemiya, T. Ohmori, T. Yamaguchi, *J. Phys. Chem. A* **104**, 336 (2000).
26. K. Miyakawa, F. Sakamoto, R. Yoshida, E. Kokufuta, T. Yamaguchi, *Phys. Rev. E* **62**, 793 (2000).
27. S. K. Scott, *Oscillations, Waves, and Chaos in Chemical Kinetics* (Oxford Univ. Press, New York, 1994).

28. This work was supported by the NSF.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/798/DC1
Fig. S1

12 July 2006; accepted 22 September 2006
10.1126/science.1132412

Renewable Hydrogen from Nonvolatile Fuels by Reactive Flash Volatilization

J. R. Salge, B. J. Dreyer, P. J. Dauenhauer, L. D. Schmidt

Droplets of nonvolatile fuels such as soy oil and glucose-water solutions can be flash evaporated by catalytic partial oxidation to produce hydrogen in high yields with a total time in the reactor of less than 50 milliseconds. Pyrolysis, coupled with catalytic oxidation of the fuels and their fragments upon impact with a hot rhodium-cerium catalyst surface, avoids the formation of deactivating carbon layers on the catalyst. The catalytic reactions of these products generate approximately 1 megawatt of heat per square meter, which maintains the catalyst surface above 800°C at high drop impact rates. At these temperatures, heavy fuels can be catalytically transformed directly into hydrogen, carbon monoxide, and other small molecules in very short contact times without the formation of carbon.

The direct conversion of nonvolatile hydrocarbons into syngas (a mixture of H_2 and CO) and chemicals is an important process for using renewable fuels such as vegetable oils and liquids produced by hydrolysis or pyrolysis of biomass (*1*). Hydrogen is needed

for fuel cells and for onboard combustion in vehicles for enhanced performance and reduced emissions, and syngas is used for the production of synthetic liquid fuels, chemicals, and fertilizers.

The conversion of gaseous and volatile fuels to H_2 is possible through pyrolysis (*1*), steam reforming (*2*), and partial oxidation (*3*–*5*), with or without catalysts. However, the direct processing of nonvolatile fuels such as vegetable oils, residual petroleum fuels, and liquid and solid biomass is more complicated because of their tendency to form solid carbon that interferes with process equipment and rapidly plugs pores in heterogeneous catalysts. Such heavy fuels decompose chemically before evaporation to form hydrogen, olefins, aromatics, and solid carbon.

Flash pyrolysis (reaction times typically 1 s) of heavy liquids and solid biomass has been shown (*1*) to produce primarily gases (syngas) and volatile liquids (bio-oils). Reaction times in these processes are limited by heat transfer into biomass particles to decompose reactants. Additionally, at least ~10% of the reactant biomass is reported to form a solid char that must be separated and removed. Nonvolatile solid biomass pellets have been shown to volatilize without the formation of carbon when exposed to very high heat fluxes (~10^6 W/m^2) (*6*) of focused radiation from a xenon lamp. We find that similar high heat fluxes can also occur by catalytic partial oxidation of biomass to H_2 and CO, providing a comparable steady-state environment where nonvolatile biomass can be decomposed without the production of carbon.

We showed that, by flash heating small drops of heavy fuels in the presence of small amounts of O_2 impinging on a catalyst-coated ceramic foam maintained at ~800°C by the reaction, it is possible to achieve steady-state operation with refined soy oil, biodiesel (the volatile methyl ester of soy oil), and sugar-water solutions with no external heat supplied. This process produces ~70% selectivity to H_2 with >99% conversion of the fuel. Carbon formation does not occur because the presence of O_2 produces rapid oxidation of decomposition products, and the resulting heat of reaction maintains a surface temperature of 800° to 1000°C that prevents quenching of the process that would lead to rapid carbon formation.

The reactor, sketched in Fig. 1, is similar to those described previously (*5*) and uses an automotive fuel injector to spray ~400-μm-diameter drops onto a catalyst foam containing Rh-Ce catalyst particles at typically 2.5% by weight of each component. We placed the catalyst ~2 cm from the fuel injector so that the cold drops impinged directly on the front face of the catalyst. Air flowed around the fuel injector to provide a uniform flow field and to optimize mixing with the gaseous products. Air and fuel enter at 20°C; no external heating was needed.

Because it is essential to have the liquid impinge on a very hot surface to prevent carbon formation, the process was initiated by flowing air mixed with CH_4 over the catalyst for light-off to heat the front face of the catalyst to ~1000°C. Then the fuel injector flow was started, and the CH_4 was reduced or shut off to maintain a constant C/O ratio (the ratio of C atoms in the fuel to the O atoms in air) of ~1.0, which is the optimum stoichiometric ratio for the formation of syngas.

Results for refined soy oil (the triglyceride of C_{18} and C_{16} fatty acids) as fuel are shown in Fig. 1A for conversion and temperature and in Fig. 1B for selectivities to H_2 and all carbon-containing products. The temperature 10 mm downstream from the front face, which is generally cooler than the front face where oxidation reactions occur, decreased from 1100°C (glowing bright orange) to 800°C (dull red) as C/O increased from 0.8 to 1.2. The lower C/O limit is set by thermal destruction of the catalyst (sintering of Rh), and the higher C/O limit is set by carbon formation that shuts down the process.

We have run this process for more than 20 hours on a given catalyst and repeated it on several catalysts. In all cases, no deactivation was observed (<2% changes in conversion or selectivities over 20 hours) as long as the catalyst was not overheated. A higher C/O ratio caused deactivation through carbon formation in the catalyst, but the activity could be restored quickly by decreasing C/O to burn off this carbon or by adding CH_4.

Results for a similar experiment with biodiesel instead of soy oil are shown in Fig. 2. Biodiesel (the methyl ester of the fatty acids from the triglyceride ester made by transesterification of soy oil) boils without decomposition at >300°C, so vaporization of biodiesel without carbon formation is possible. We showed previously (*4*) that biodiesel could be converted into H_2 and CO or into olefins in a similar reactor system where the biodiesel was vaporized by heating the walls of the reactor above the catalyst to 300°C, and a heat shield–mixer was inserted between the vaporized fuel and the catalyst. In the present experiment, no external heat was added, so fuel and air were fed at room temperature. Figure 2 shows that conversions and selectivities with biodiesel are similar to those seen when using nonvolatile soy oil directly.

In another experiment, we fed a glucose-water solution through the fuel injector. We dissolved 35% by weight glucose in water. (The solubility of glucose in water at 20°C is 38% by weight.) We were unable to maintain steady-state operation without adding CH_4 along with the sugar, and the lowest CH_4 level we were able to maintain at C/O = 0.8 was three carbon atoms from methane per carbon atom from glucose. Again, performance was comparable to that seen with soy oil or biodiesel, with >99% glucose conversion and ~70% H_2 selectivity.

These are preliminary experiments in which little optimization of catalyst or conditions was attempted. Preheat in the volatilization zone or in the fuel injector should increase stability and performance, and smaller drop sizes should also increase the range of operation.

This process produces no more than a few percent of higher hydrocarbons (primarily ethylene and propylene) for any of these fuels at C/O = 0.8, so these products are suitable for use in a fuel cell with minimal cleanup. We observed up to 20% carbon atom selectivity to ethylene and propylene at C/O = 1.2, and operation at higher C/O ratios should produce much higher yields of olefins, comparable to the

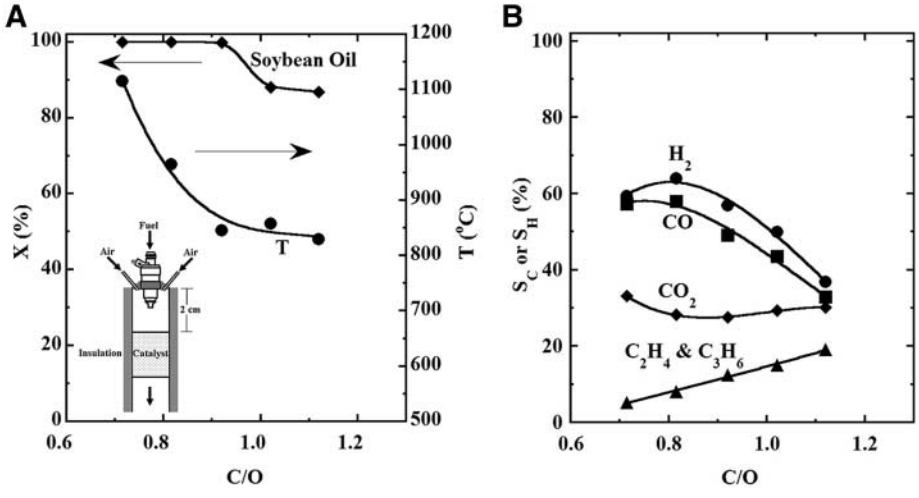


Fig. 1. Soy oil (**A**) conversion and temperature and (**B**) selectivities to H_2 and carbon products by reactive flash volatilization. S_C and S_H are defined as (C or H atoms in product)/(C or H atoms in converted fuel). All gases and liquids enter the reactor at room temperature, and no process heat is added. The inset shows the reactor configuration, in which liquid drops of soy oil from an automotive fuel injector strike a Rh-Ce catalyst on an alumina foam to vaporize the drops in ~1 ms without carbon formation.

Department of Chemical Engineering and Material Science, University of Minnesota, 421 Washington Avenue SE, Minneapolis, MN 55455, USA.

80% total olefins reported for conventional catalytic partial oxidation of biodiesel (*4*) and hexadecane (*5*). Operation at higher C/O ratios would require the adjustment of preheat and flow conditions and perhaps the use of sacrificial CH_4 to avoid carbon formation. A patent describing a similar experiment has been reported, in which heavy petroleum was co-fed with light hydrocarbons or fed periodically in order to produce ethylene and higher olefins (*7*).

We next consider the processes by which nonvolatile liquids can be converted to H_2 and other small molecules without carbon formation. When a drop of a volatile liquid hits a hot surface, the vaporization at the interface can be fast enough that a gaseous layer is formed, which insulates the drop from the surface. This regime is called "film boiling" or "water droplets on a hot frying pan," as sketched in Fig. 3A, and the heat transfer and mechanisms of this process have been considered extensively (*8*–*11*). Nonreactive drop breakup upon impact has been studied (*12*), as has nonreactive drop impact upon heated porous surfaces (*13*).

The situation in our experiment is considerably different than conventional film boiling. Without chemical reaction, the surface would cool rapidly by heat transfer and boiling of the liquid. Instead, the process relies on the chemical reaction of the volatile products (the parent liquid as well as H_2 and other smaller fragments) to continuously maintain the surface temperature high enough to sustain steady-state impact and decomposition of drops. The possible configuration for a drop with reaction occurring on a hot catalyst surface is sketched in Fig. 3B. The process probably relies on reaction in the gaseous layer between the drop and the hot catalyst to continuously generate gaseous products and heat, and the process can continue until all fuel is volatilized, either in the initial drop or in successive smaller drops that form from the primary drop.

The sequence of surface and homogeneous reaction steps in reactive flash volatilization is unknown. We suspect that the initial reaction step is partial vaporization and pyrolysis upon impact of the cold drop with the hot surface. Rebound then allows oxidation reactions, and film boiling permits gas and surface reactions. Once the drop or its fragments enter the ceramic foam, reactive decomposition should be very rapid. However, the amounts of reaction in each of these stages are unknown.

The overall process is extremely complicated because of possible drop breakup dynamics and the structure of the porous catalyst surface on which the process occurs. Some of these issues are sketched in Fig. 3, C and D. The velocity of the drop above the surface is calculated to be ~1 m/s, so the impact of the drop on the surface could promote rapid breakup into smaller drops by momentum transfer. The initial impact probably involves the cold liquid making direct contact with the surface. These events are estimated to involve microsecond times, where gradients and sequences of events are extremely large and difficult to predict. Previous studies have also noted the difficulties in accurately describing the dynamics of droplets that simultaneously boil and crack when striking a hot surface (*11*) or that strike a hot porous substrate (*13*).

This process appears to be general for any fuel, because the hot catalyst surface will pyrolyze and oxidize the liquid at the interface into products that are easily oxidized. Heavy fuels such as residual petroleum fractions, yellow grease (used cooking oil), and crude soy oil should be processable with little or no pretreatment. At the observed reactor temperatures of 800° to 1000°C, the presence of small amounts of impurities such as phosphorous, potassium, nitrogen, and other small organics present in renewable feedstocks will probably not be detrimental to the catalyst because they are volatile. Varying amounts of moisture within feedstocks should actually improve the H_2 yield and further suppress carbon formation. The process should also be scalable over a wide range of capacities (*14*).

Many technologies require fast drop volatilization without carbon formation. Diesel engines require rapid combustion of nonvolatile fuels, but the impact of drops on walls is generally avoided to prevent coke formation. New engine technologies could use drop volatilization at a catalytic surface to improve diesel combustion and reduce pollutants. Heavy oils such as residual petroleum fractions and biomass-derived liquids can be pyrolyzed and combusted in fixed or fluidized beds, but this generally involves a reducing zone where carbon forms, followed by an oxidizing zone where the carbon is burned off. A single-

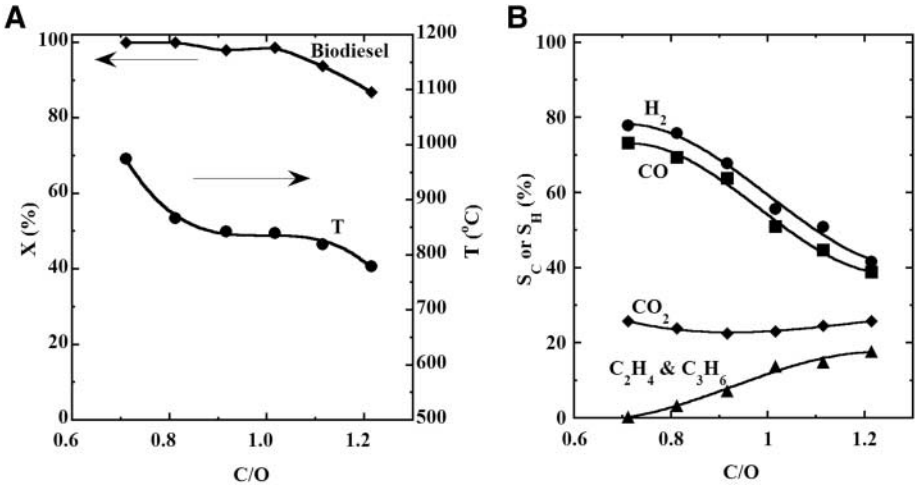


Fig. 2. Biodiesel (**A**) conversion and temperature and (**B**) selectivities to H_2 and carbon products by reactive flash volatilization. Conversion is high and selectivities are comparable to those seen with soy oil.

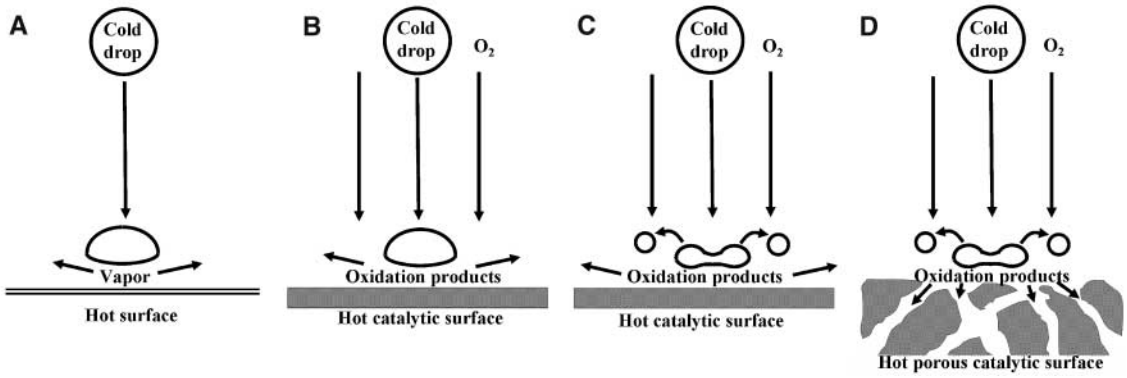


Fig. 3. Sketches of possible configurations of (**A**) conventional film boiling of a volatile drop on a hot surface, (**B**) reactive drop volatilization on a hot catalyst surface, (**C**) drop impingement and breakup on a hot catalytic surface, and (**D**) drop impingement and breakup on a hot catalytic porous surface.

zone catalytic process would be much smaller and simpler, and the use of catalysts would allow tuning of selectivities that is not possible with flame combustors. Catalytic processes also eliminate or strongly reduce pollution associated with flame combustors.

Reactive drop volatilization appears to be a simple and readily adaptable method to convert nonvolatile fuels into H_2 or chemicals for large as well as small scales of production, such as onboard vehicle reforming. It allows the intensification of the process into millisecond time scales and suggests that the conversion of other nonvolatile biomass mixtures such as emulsions, slurries, and powders is possible. The process also requires further experiments, long-term evaluation, and modeling to optimize catalyst performance and determine the exact mechanisms of reactive flash volatilization.

References and Notes

1. A. V. Bridgwater, *Chem. Eng. J.* **91**, 87 (2003).
2. L. Garcia, R. French, S. Czernik, E. Chornet, *Appl. Catal. A* **201**, 225 (2000).
3. G. A. Deluga, J. R. Salge, L. D. Schmidt, X. E. Verykios, *Science* **303**, 993 (2004).
4. R. Subramanian, L. D. Schmidt, *Angew. Chem. Int. Ed.* **44**, 302 (2005).
5. J. J. Krummenacher, K. N. West, L. D. Schmidt, *J. Catal.* **215**, 332 (2003).
6. O. Boutin, M. Ferrer, J. Lede, *Chem. Eng. Sci.* **57**, 15 (2002).
7. D. C. Griffiths, K. W. Palmer, I. A. B. Reid, U.S. Patent 5,663,473 (1997).
8. S. Deb, S.-C. Yao, *Int. J. Heat Mass Transfer* **32**, 2099 (1989).
9. L. H. J. Wachters, N. A. J. Westerling, *Chem. Eng. Sci.* **21**, 1047 (1966).
10. Y. Ge, L.-S. Fan, *Phys. Fluids* **17**, 027104 (2005).
11. B. S. Gottfried, C. J. Lee, K. J. Bell, *Int. J. Heat Mass Transfer* **9**, 1167 (1966).
12. M. Bussmann, S. Chandra, J. Mostaghimi, *Phys. Fluids* **12**, 3121 (2000).
13. S. Chandra, C. T. Avedisian, *Int. J. Heat Mass Transfer* **35**, 2377 (1992).

14. Our system processes approximately 0.6 kg/day of fuel using 150 mg of Rh, and a catalyst disc 5 cm in diameter would process ~5.2 kg/day under identical conditions. We had to use small single-orifice automotive fuel injectors to obtain sufficiently low flows, but larger systems could use larger multiport injectors, multiple injectors, or different methods for uniform drop formation over the entire catalyst surface.

15. This research was supported by grants from the U.S. Department of Energy (DE-FG02-88ERB878) and the Initiative for Renewable Energy and the Environment at the University of Minnesota (LG-MC1-2005).

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/801/DC1
Materials and Methods
Figs. S1 and S2
References

13 June 2006; accepted 20 September 2006
10.1126/science.1131244

Coherent Electronic Fringe Structure in Incommensurate Silver-Silicon Quantum Wells

N. J. Speer, S.-J. Tang, T. Miller, T.-C. Chiang*

Atomically uniform silver films grown on highly doped n-type Si(111) substrates show fine-structured electronic fringes near the silicon valence band edge as observed by angle-resolved photoemission. No such fringes are observed for silver films grown on lightly doped n-type substrates or p-type substrates, although all cases exhibited the usual quantum-well states corresponding to electron confinement in the film. The fringes correspond to electronic states extending over the silver film as a quantum well and reaching into the silicon substrate as a quantum slope, with the two parts coherently coupled through an incommensurate interface structure.

Electronic effects in thin films and at interfaces are at the heart of modern solid state electronic technology, and as device dimensions shrink toward the nanoscale, quantum coherence and interference phenomena become increasingly important. For example, a thin film of Ag on Si, a prototypical metal-semiconductor system with an incommensurate interface, can support quantum-well states caused by electron confinement in the film (*1*–*5*). We now report on electronic fringes in atomically uniform Ag films grown on Si(111) that exhibit much finer structures than previously observed. The fringes of dispersion curves closely track the Si substrate valence band edge in energy-momentum space and bear a striking visual resemblance to ripples associated with diffraction of waves at a boundary. Although both p- and n-doped Si substrates were studied, only highly doped n-type Si substrates exhibit such fringes, which suggests that the effects must be related to the electronic potential variation in the substrate. The technique we used, angle-resolved photoemission, has a very short probing depth of just two to three atomic layers. Nevertheless, the wave function sampled by photoemission propagates through the film and into the depletion region in the Si to a depth up to tens of atomic layers. A long electron coherence length gives rise to quantum interference, despite the large lattice mismatch between Ag and Si and the incommensurate interface configuration. The resulting electronic structure is characteristic of the combined Ag and Si system, and an analysis of the wave mechanics yields results matching closely the experimental observation.

In our experiment, the Si(111) substrates were flashed to create a clean 7 by 7 reconstructed surface. Silver films were deposited at a substrate temperature of 50 K; these were subsequently annealed to near room temperature and cooled back to 50 K for photoemission measurements. Angle-resolved photoemission data taken from Ag films of a thickness of $N = 8$ monolayers (ML) deposited on n-type Si substrates are shown in Fig. 1A for a doping level of $n = 2 \times 10^{15}/\text{cm}^3$ (lightly doped) and Fig. 1B for $n = 5 \times 10^{18}/\text{cm}^3$ (highly doped). Under our experimental conditions, Ag(111) films grow on Si(111) with an unstrained bulk Ag lattice constant, and the $\overline{\Gamma}\,\overline{K}$ direction (*6*), corresponding to the $[\bar{1}10]$ direction in the bulk, is aligned with the same direction in the substrate. The emission angle refers to the polar angle relative to the surface normal, and the detection plane is oriented along the $\overline{\Gamma}\,\overline{K}$ direction. The results from the lightly doped sample (Fig. 1A) show a surface state (SS) similar to that of bulk Ag(111) (*7*) and a set of quantum-well states (or subbands) labeled by the quantum number $\nu = 1$–3. These subbands exhibit approximately parabolic dispersion relations, but with notable "kinks" near the top Si valence band edge caused by a hybridization interaction (*8*–*10*). An example of such a kink is indicated by an arrow. The data for the highly doped sample (Fig. 1B) show similar features and, additionally, fringes near the Si valence band edge. An enlarged view of the region in Fig. 1B contained within the rectangular box shows details of the fringes (Fig. 1C).

The film thickness of $N = 8$ ML quoted for the data is an exact thickness. A counting of the atomic layers over a wide range of film coverage leads to an absolute determination of the film thickness (*8*). In Fig. 2, A to C, we show data for Ag films with coverages of 11, 11.5, and 12 ML, respectively, on highly doped n-type Si. Similar fringe structures were observed for these three coverages, but the quantum-well subbands in Fig. 2B additionally show splittings at places where they are narrow in energy. An example lies in the circled area. The splittings are caused by the presence of two thicknesses, 11 and 12 ML, in the film. This atomic layer resolution allows us to follow the completion of each atomic layer as we gradually increase the film thickness.

Because the fringes depend on the substrate doping, the substrate potential must play an important role in the observed effects. Furthermore,

Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801–3080, USA, and Frederick Seitz Materials Research Laboratory, University of Illinois at Urbana-Champaign, 104 South Goodwin Avenue, Urbana, IL 61801–2902, USA.

*To whom correspondence should be addressed. E-mail: chiang@mrl.uiuc.edu

each fringe is smoothly connected to a quantum-well subband, as seen in Fig. 1. Thus, the fringes can be viewed as a bundle of quantum-well subbands that pile up near the Si valence band edge. This behavior is not expected based on the standard model that treats the Ag film as a quantum well, with the Si substrate simply providing a boundary that reflects the electrons (*1*). To explain the results, we show in Fig. 3 a plot of the electronic potential for a film of 8 ML at in-plane wave vector $k_x = 0.22$ Å^{-1} (corresponding to a polar emission angle of ~6°). The Si substrate has a gap in which the Fermi level E_F lies. At the Ag-Si interface, E_F is pinned near midgap (*11*). Band bending in the depletion region of Si gives rise to an approximately linear depth dependence of the valence band edge, as indicated in Fig. 3 (*12*). Propagating electronic states in Si exist only below this edge. There is no gap in Ag at this k_x, and all Ag states below E_F are propagating in nature.

The electronic potentials in the Ag film and the Si substrate together create a set of confined states. To solve the wave mechanical problem, we used the effective mass approximation (*13*, *14*). The wave function in the Ag film is given by:

$$\psi \propto \exp(ik_x x)\cos\left(\frac{\pi z}{t}\right)\sin[k_z(z - Nt - \Delta)] \quad (1)$$

where k_z is the momentum component perpendicular to the surface measured from the zone boundary, t is the Ag monolayer thickness, and Δ is a charge spillage parameter (*15*), which defines an effective boundary at the Ag-vacuum interface. Here, $z = 0$ is the Ag-Si boundary, and the wave function vanishes at $z = Nt + \Delta$. The three factors on the right-hand side of Eq. 1 can be identified as the in-plane wave function, the band-edge wave function, and the envelope function, respectively. The wave function in Si is similarly given by a product of three factors:

$$\psi \propto \exp(ik_x x)\cos\left(\frac{2\pi}{a}\sqrt{3}z\right) \times \mathrm{Ai}\left(\left(\frac{2m^*}{F^2\hbar^2}\right)^{\frac{1}{3}}[U(z) - E]\right) \quad (2)$$

where a is the lattice constant of Si, U is the linear potential caused by the band bending, F is the slope of this potential, E is the energy of the state, m^* is the Si effective mass along the z direction, and Ai denotes the Airy function. The Airy function is a solution of the Schrödinger equation with a linear potential and is the form of the envelope function in Si (*14*).

The band parameters of Ag and Si were taken from an empirical tight binding calculation (*16*, *17*). The potential and its slope in the Si substrate were calculated from the dopant density. The parameter Δ was taken to be 0.4 Å (0.17 ML); the calculated results were not very sensitive to this choice. The only free parameter in the model was the energy of the Si band edge at the Ag-Si interface. The wave functions in Eqs. 1 and 2 were matched, and the Si band edge was varied for a best fit to the observed dispersion relations. The photoemission data for the 8-ML film with the Si band edge indicated are shown in Fig. 4A, and Fig. 4B shows the same data with the calculated dispersion relations superimposed. The agreement is very good for the fringe spacings and dispersions.

The wave functions for the first five states, counting from E_F, are shown in Fig. 3 for $k_x = 0.22$ Å^{-1}. The first one ($\nu = 1$) lies completely within the Si band gap. Its wave function is confined within the Ag film and decays rapidly inside the Si substrate. The other four states, at lower energies, penetrate into the Si depletion region to various depths. With each increment of ν within the group of four states, the energy decreases by ~70 meV, and the wave function penetrates deeper by ~10 Å. The wave functions remain fairly similar within the Ag film. The relatively shallow slope of the potential within the Si causes the electronic states with different ν's to pile up near the Si band edge, giving rise to the closely spaced fringes. As k_x increases, the Si band edge moves down, and more states become confined within the Ag film, as seen in Fig. 4. States with higher ν's should also be present in the data, but they become fainter for two reasons. The probing depth of photoemission is only a few angstroms. Wave functions extending deeper into the substrate have less weight in the photoemission probing depth, and the emission intensity decreases correspondingly. Also, a finite electron coherence length in the system can limit the range over which discrete states can be clearly observed (*18*).

For lightly n-doped Si substrates, the slope of the potential in Si is essentially zero. No fringes are expected, and none are observed. Likewise, no fringes are expected or observed for p-doped substrates, because the direction of band bending

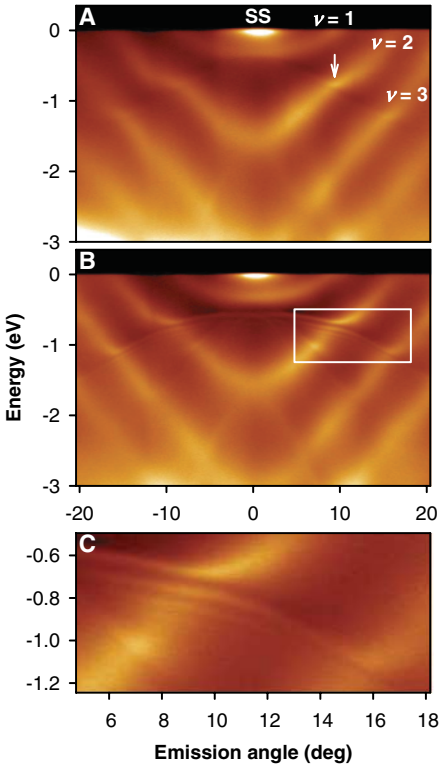


Fig. 1. Angle-resolved photoemission data for 8 ML of Ag grown on (**A**) lightly doped n-type Si and (**B**) highly doped n-type Si. (**C**) is an enlarged view of the region contained within the rectangular box in (B). The photon energy used was 22 eV.

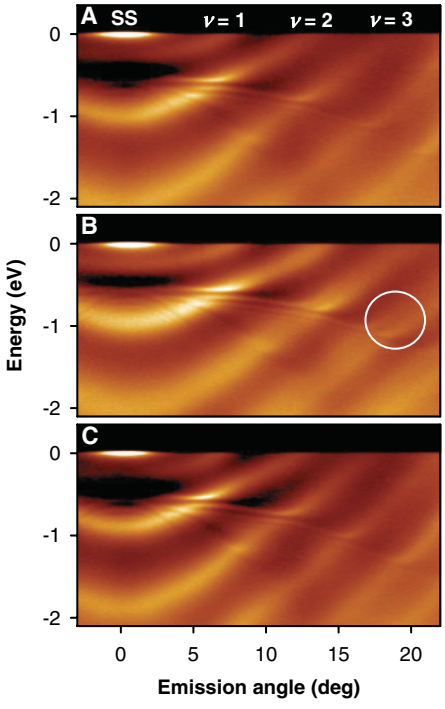


Fig. 2. Photoemission data for (**A**) 11-, (**B**) 11.5-, and (**C**) 12-ML Ag films grown on highly doped Si. The data in (B) show band splitting due to the presence of two thicknesses. An example is indicated by the circle.

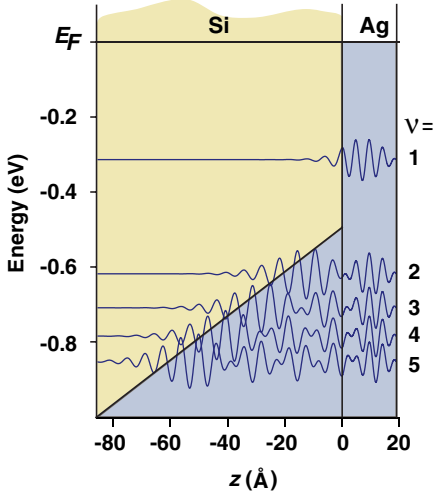


Fig. 3. Plot of the highly doped Si valence band and first five wave functions at $k_x = 0.22$ Å^{-1} for a film thickness of 8 ML.

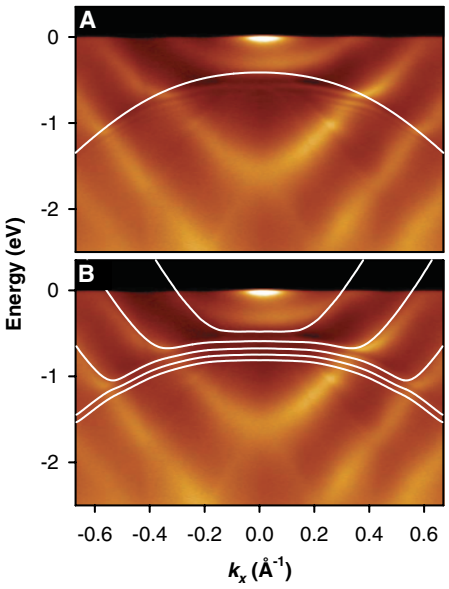


Fig. 4. Photoemission data for 8-ML Ag on highly doped Si(111). The horizontal axis has been converted from emission angle to k_x. (**A**) The curve indicates the position of the Si valence band edge. (**B**) The curves show the calculated energy dispersion relations.

is opposite to that of the n-type substrates. We note that the $\nu = 2$ and 3 states in Fig. 1A for the lightly doped sample, instead of bending over to form fringes, simply continue into the continuum region of Si, with a kink as the dispersion curve crosses the Si band edge. These states within the Si band continuum are actually quantum-well resonances, as they are not fully confined. The Ag-Si boundary causes partial reflection; the resulting interference effect gives rise to broadened, quasi-discrete states. Such quantum-well resonances are also present in Fig. 1B for the highly doped sample at energies beyond the range of band bending (or confinement).

The Ag films and the Si substrates are lattice mismatched and incommensurate. Nevertheless, the wave functions given in Eqs. 1 and 2 are symmetry compatible and can be matched over the entire interface plane. The resulting state is coherent throughout the entire system. The combination of a quantum well (Ag film) and a quantum slope (Si substrate) yields a rich electronic structure. An important conclusion drawn from the present study is that coherent wave function engineering, as is traditionally carried out in lattice-matched epitaxial systems, is entirely possible for incommensurate systems, which can substantially broaden the selection of materials useful for coherent device architecture.

References and Notes

1. T.-C. Chiang, *Surf. Sci. Rep.* **39**, 181 (2000).
2. A. L. Wachs, A. P. Shapiro, T. C. Hsieh, T. C. Chiang, *Phys. Rev. B* **33**, 1460 (1986).
3. S.-Å. Lindgren, L.Walldén, *Handbook of Surface Science*, Vol. 2, *Electronic Structure*, edited by S. Holloway, N. V. Richardson, K. Horn, M. Scheffler (Elsevier, Amsterdam, 2000).
4. M. Milun, P. Pervan, D. P. Woodruff, *Rep. Prog. Phys.* **65**, 99 (2002).
5. F. J. Himpsel, J. E. Ortega, G. J. Mankey, R. F. Willis, *Adv. Phys.* **47**, 511 (1998).
6. S.Hüfner, *Photoelectron Spectroscopy* (Springer-Verlag, New York, ed. 2, 1996).
7. G. Nicolay, F. Reinert, S.Hüfner, P. Blaha, *Phys. Rev. B* **65**, 033407 (2002).
8. S.-J. Tang, L. Basile, T. Miller, T.-C. Chiang, *Phys. Rev. Lett.* **93**, 216804 (2004).
9. I. Matsuda, T. Ohta, H. W. Yeom, *Phys. Rev. B* **65**, 085327 (2002).
10. L. Aballe, C. Rogero, P. Kratzer, S. Gokhale, K. Horn, *Phys. Rev. Lett.* **87**, 156801 (2001).
11. S. Kurtin, T. C. McGill, C. A. Mead, *Phys. Rev. Lett.* **22**, 1433 (1969).
12. S. N. Takeda, N. Higashi, H. Daimon, *Phys. Rev. Lett.* **94**, 037401 (2005).
13. O. Madelung, *Introduction to Solid State Theory* (Springer-Verlag, New York, 1978).
14. J. H. Davies, *The Physics of Low-Dimensional Semiconductors: An Introduction* (Cambridge, New York, 1998).
15. P. Czoschke, H. Hong, L. Basile, T.-C. Chiang, *Phys. Rev. B* **72**, 035305 (2005).
16. D. A. Papaconstantopoulos, *Handbook of the Band Structure of Elemental Solids* (Plenum, New York, 1986).
17. Program *Static* developed by the Center for Computational Materials Science, Naval Research Laboratory (http://cst-www.nrl.navy.mil/bind/static/index.html).
18. J. J. Paggel, T. Miller, T.-C. Chiang, *Science* **283**, 1709 (1999).
19. Supported by the U.S. Department of Energy (grant DEFG02-91ER45439). We acknowledge the Petroleum Research Fund, administered by the American Chemical Society, and the U.S. National Science Foundation (grant DMR-05-03323) for partial support of personnel and the beamline facilities at the Synchrotron Radiation Center, where the photoemission work was performed. The Synchrotron Radiation Center is supported by the U.S. National Science Foundation (grant DMR-05-37588).

24 July 2006; accepted 8 September 2006
10.1126/science.1132941

Solar Nebula Heterogeneity in p-Process Samarium and Neodymium Isotopes

Rasmus Andreasen and Mukul Sharma*

Bulk carbonaceous chondrites display a deficit of ~100 parts per million (ppm) in ^{144}Sm with respect to other meteorites and terrestrial standards, leading to a decrease in their ^{142}Nd/^{144}Nd ratios by ~11 ppm. The data require that samarium and neodymium isotopes produced by the p process associated with photodisintegration reactions in supernovae were heterogeneously distributed in the solar nebula. Other samarium and neodymium isotopes produced by rapid neutron capture (r process) in supernovae and by slow neutron capture (s process) in red giants were homogeneously distributed. The supernovae sources supplying the p- and r-process nuclides to the solar nebula were thus disconnected or only weakly connected.

Planetary crust-mantle differentiation leads to fractionation of Sm from Nd producing variations in the Nd isotopic compositions from the radioactive decay of ^{146}Sm to ^{142}Nd [half-life of ^{146}Sm = 103 million years (My)] and of ^{147}Sm to ^{143}Nd [half-life of ^{147}Sm = 106 billion years (Gy)]. Deviations in the Nd isotopic compositions of crust-mantle reservoirs from the bulk planetary values reflect the timing and magnitude of planetary evolution. A fundamental assumption in the interpretation of isotope records is that the Nd isotopic composition of bulk planets is identical to that of chondritic meteorites. Recent results show that the ^{142}Nd/^{144}Nd ratios of chondrites and the Moon (*1*, *2*) are ~20 μ (μ = parts in 10^6) lower than the terrestrial upper mantle, which suggests that Earth contains a complementary reservoir with a low Sm/Nd ratio that resulted from a global differentiation event >4.53 billion years ago (Ga). The Nd isotopic compositions of chondrites are, however, somewhat heterogeneous, displaying –5 to –40 μ variations in ^{142}Nd (*1*) without the correlated variations in ^{143}Nd, as expected if there were small differences in the Sm/Nd ratio of the chondrites. Chondrites also preserve grains showing large anomalous variations in the Sm and Nd isotopes produced during nucleosynthesis [e.g., (*3*)]. These observations raise a key question: Were the nucleosynthetic contributions of Sm and Nd isotopes heterogeneously distributed in the solar nebula isotopes on a planetary (or planetesimal) scale? We sought to answer this question by examining Sm and Nd isotopes in macroscopic samples of meteorites.

Isotopes produced by the r process (rapid neutron capture) and the p process (photodissociation of nuclides) are considered to be generated in supernovae explosions; those produced by the s process (slow neutron capture) are generated in red giants [e.g., (*4*)]. Relative contributions of stellar sources of Sm and Nd isotopes (Fig. 1) indicate that coupled

Radiogenic Isotope Geochemistry Laboratory, Department of Earth Sciences, Dartmouth College, 6105 Sherman Fairchild Hall, Hanover, NH 03755, USA.

*To whom correspondence should be addressed. E-mail: mukul.sharma@dartmouth.edu

variations in $^{144}Sm/^{154}Sm$, $^{148}Sm/^{154}Sm$, and $^{152}Sm/^{154}Sm$ can be used to quantify p-, s-, and r-process contributions. Additionally, all Nd isotopes, with the exception of radiogenic ^{142}Nd and ^{143}Nd, can be used to evaluate variations in the contributions of s and r processes. Moreover, ^{149}Sm that has a large thermal neutron cross section can be used to estimate the neutron fluence and its affect on other isotopes of the Sm-Nd system.

Powdered meteorite samples weighing 0.5 to 1 g were dissolved [see (*5*) for details of analytical methods]. Sm and Nd concentrations were determined by isotope dilution using ~5% aliquots of the resulting solutions (table S1). Other larger aliquots of the sample solution were processed to separate Sm and Nd. Measurements of Sm isotopes for meteorite samples were interleaved with those of a standard, which is assumed to give the Sm isotopic composition of the terrestrial upper mantle; analyses of terrestrial basalts support this assumption (tables S2 and S3). The Sm isotopic compositions of Saint Severin [ordinary chondrite (O-chondrite)] and Juvinas (eucrite) are identical to that of terrestrial Sm (Fig. 2 and table S3). In contrast, Allende and Murchison [carbonaceous chondrites (C-chondrites)] display significant negative ^{144}Sm anomalies with respect to terrestrial Sm (Fig. 2 and table S3). No other Sm isotopes display anomalous abundances except ^{149}Sm and ^{150}Sm, which are affected by neutron capture ($^{149}Sm + {}^{1}n \rightarrow {}^{150}Sm$; table S3). Could the negative ^{144}Sm anomaly be a result of neutron-capture reactions? Large thermal neutron fluences capable of changing all Sm and Nd isotope ratios have been reported for aubrites and lunar samples [e.g., (*2*, *6*, *7*)]. Using the deficits in ^{149}Sm in the meteorites, we estimate the neutron fluences to be 1×10^{14} to 6×10^{14} n cm^{-2} (table S3), similar to those of (*8*). Calculations of neutron-capture effects show that the observed deficit in ^{144}Sm cannot be caused by exposure to thermal neutrons (*5*).

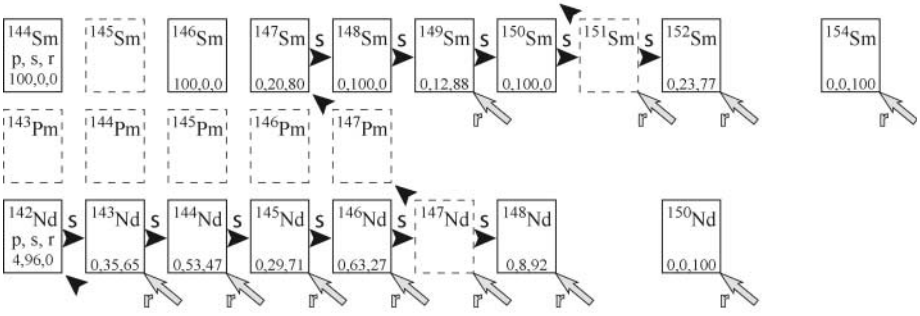


Fig. 1. Chart of nuclides showing the s-process path (arrowheads) in the vicinity of Sm and Nd. Isotopes on the proton-rich side of the path (^{144}Sm, ^{146}Sm) are bypassed by the s process and require a separate p process. Isotopes on the neutron-rich side of the path (^{154}Sm, ^{150}Nd) are also bypassed by the s process and are produced by rapid neutron capture (r process, arrows). Values separated by commas are the relative contributions of the p, s, and r processes for stable and long-lived isotopes (*10*, *11*).

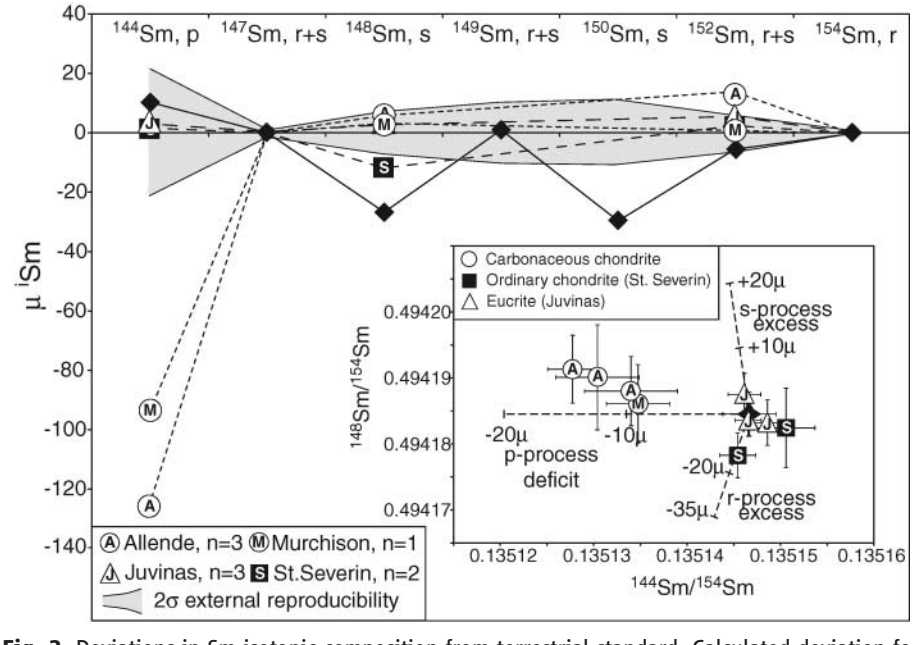


Fig. 2. Deviations in Sm isotopic composition from terrestrial standard. Calculated deviation for terrestrial Sm minus 10 μ s-process material (*10*) is shown as a solid line. The uncertainties in the calculations are about the same size as the symbols. ^{149}Sm and ^{150}Sm data are omitted from the meteorite samples, as they are affected by thermal neutron flux. Inset is a three-isotope plot of $^{144}Sm/^{154}Sm$ and $^{148}Sm/^{154}Sm$ showing calculated trajectories for r- and s-process excesses and p-process deficit (*10*). The μ-values given in the inset correspond to the resulting changes in $^{142}Nd/^{144}Nd$ ratio.

How would small variations in nucleosynthetic contributions influence the measured Sm isotopic ratios, and what would be the effect on the measured Nd isotopic ratios? The $^{148}Sm/^{154}Sm$ ratio, which is very sensitive to s- and r-process variations, is quite uniform among the meteorites (Fig. 2), indicating that the solar nebula was homogeneous with respect to s- and r-process contributions to Sm. In contrast, the observed deficit in ^{144}Sm in C-chondrites is intriguing. A p-deficit, an s-excess, or an r-excess will lower the $^{144}Sm/^{154}Sm$ ratio (Fig. 2, inset). The most straightforward explanation for the negative ^{144}Sm anomalies in Murchison and Allende is a p-process deficit relative to Saint Severin, Juvinas, and terrestrial Sm. This indicates that the p-process isotopes were heterogeneously distributed in the solar nebula.

Because a change in the p-process contribution affects the $^{142}Nd/^{144}Nd$ ratio by primary production of ^{142}Nd and α-decay of ^{146}Sm (Fig. 1), it follows that the observed deficit of ^{144}Sm in C-chondrites should have a corresponding deficit of ^{142}Nd. The average ^{144}Sm anomaly for Allende and Murchison is –118 ± 38 μ, which corresponds to a lowering of the $^{146}Sm/^{144}Sm$ ratio in these meteorites by ~2% with respect to the solar system initial value of 0.008 ± 0.001 (*9*). The calculation of the $^{142}Nd/^{144}Nd$ ratio depends on the p-process ^{146}Sm, ^{144}Sm, and ^{142}Nd yields of the stellar models. With currently accepted stellar models (*10*, *11*), the p-process ^{146}Sm and ^{142}Nd deficits in C-chondrites would change their $^{142}Nd/^{144}Nd$ ratios by -11^{+7}_{-13} μ relative to O-chondrites and eucrites, which display no anomaly in Sm (Fig. 2) (*5*).

To evaluate this issue further, we obtained Nd isotopic ratios on the same sample solutions for which the Sm data were acquired (*5*) (tables S4 and S6). The difference in ^{142}Nd anomalies between C-chondrites and O-chondrites is about –10 μ, which is the same as the calculated difference based on the p-process deficit in ^{144}Sm (Fig. 2, inset). The nearly invariant $^{145}Nd/^{144}Nd$ and $^{148}Sm/^{154}Sm$ ratios (Figs. 2 and 3) provide an unambiguous evidence of s- and r-isotope homogeneity within our current level of resolution. This is in contrast to observed excesses in ^{137}Ba and ^{138}Ba for C-chondrites over terrestrial standards that have

been attributed to r-process excesses (*12*); this discrepancy suggests that a different mechanism may be needed to produce the Ba isotopes. Also, the $^{144}Sm/^{154}Sm$ ratio indicates that the solar nebula was heterogeneous for the p-process Sm and Nd isotopes at a level of ~1 ε (ε = parts in 10^4); the latter inference is attested by variations in $^{142}Nd/^{144}Nd$ ratios. These observations imply that the sources of p- and r-process nuclides presumed to be supernovae (SNe) were decoupled or only occasionally coupled.

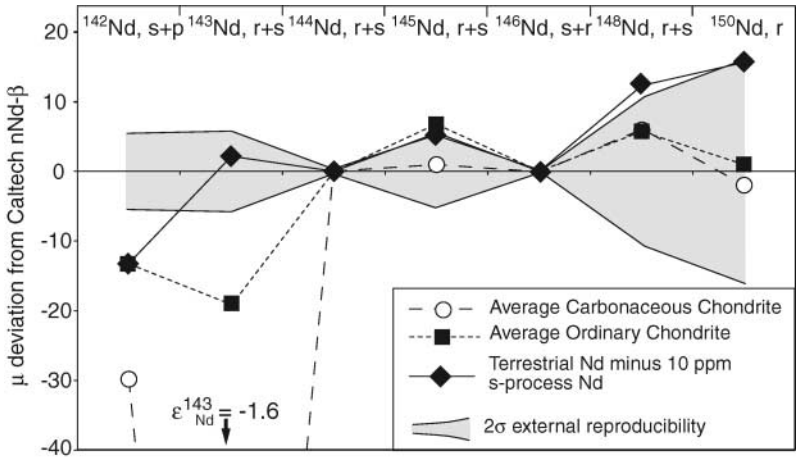


Fig. 3. Deviations in Nd isotopic composition from terrestrial standard for average C-chondrite and average O-chondrite [data from this study and (*1*)], along with calculated deviation for terrestrial Nd minus 10 μ s-process Nd (*11*). The uncertainties in the calculations are about the same size as the symbols.

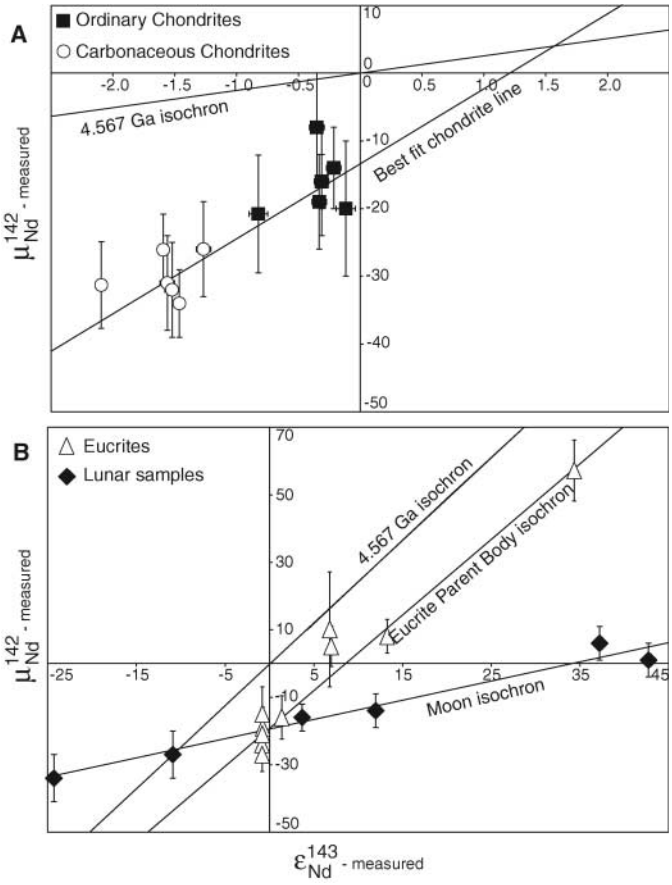


Fig. 4. (**A**) Measured μ^{142}_{Nd} and ε^{143}_{Nd} for C-chondrites (circles) and O-chondrites (squares) [this study; (*1*)]. The 4.567-Ga isochron is calculated using $\varepsilon^{143}_{Nd} \approx Q_{143} f_{Sm/Nd} t$ and $\mu^{142}_{Nd} \approx Q'_{142} f_{Sm/Nd} (^{146}Sm/^{144}Sm)_t \exp(-\lambda_{146} t)$ with Q_{143} = 25.09 Gy^{-1}, Q'_{142} = 35,400 Gy^{-1}, $^{146}Sm/^{144}Sm_t$ = 0.008, t = 4.567 Gy, λ_{146} = 6.73 Gy^{-1}, and $f_{Sm/Nd} = [(^{147}Sm/^{144}Nd_{sample})/(^{147}Sm/^{144}Nd_{CHUR})] - 1$. ε^{143}_{Nd} is 1 part in 10^4 deviation from $^{143}Nd/^{144}Nd$ = 0.512638; μ^{142}_{Nd} is 1 part in 10^6 deviation from $^{142}Nd/^{144}Nd$ = 1.141833. The Nd isotopic composition is assumed to reflect a chondritic uniform reservoir (CHUR) (*13*, *14*), although the latter ratio is measured for a laboratory Nd standard (table S4). Objects formed at the time of solar system formation with different $f_{Sm/Nd}$ values will define this isochron, assuming no subsequent disturbance. The chondrite data plot below the 4.567-Ga isochron and define a slope that gives $^{146}Sm/^{144}Sm_t = 0.035^{+12}_{-11}$, which is a factor of 3 to 5 higher than the solar system $^{146}Sm/^{144}Sm_t$ value. If the C-chondrites corrected for p-process deficit are plotted with O-chondrites, we would obtain a line consistent with the 4.567-Ga isochron but with an offset of −16 ± 4 μ, implying that the CHUR $^{142}Nd/^{144}Nd$ ratio is lower than assumed by this amount. The p-process correction is model dependent and adds considerable uncertainty to the corrected $^{142}Nd/^{144}Nd$ ratio for the C-chondrites. (**B**) Measured μ^{142}_{Nd} and ε^{143}_{Nd} for eucrites (triangles) [this study; (*1*)] and lunar samples (diamonds) (*2*). The Eucrite Parent Body and Moon isochrons give ages of 4554^{+24}_{-28} Ma and 4349^{+44}_{-63} Ma, respectively; both have intercepts of −19 ± 3 (2σ).

The ^{144}Sm deficit, combined with the coupled $^{146,147}Sm$-$^{142,143}Nd$ isotope systematics of meteorites, indicates that C-chondrites cannot be used to provide a precise estimate of the bulk planetary Nd isotopic composition (Fig. 4A) and that a better assessment is provided by O-chondrites. A weighted average $^{142}Nd/^{144}Nd$ ratio of all bulk O-chondrites analyzed so far is 16 ± 4 μ units below Nd standard (≡ terrestrial upper mantle) (table S6). That the bulk planetary $^{142}Nd/^{144}Nd$ ratio is similar to that of O-chondrites is supported by the estimated $^{142}Nd/^{144}Nd$ ratios of the Moon and Eucrite Parent Body (EPB, likely asteroid 4 Vesta), which are both 19 ± 3 μ units below that of the terrestrial upper mantle (Fig. 4B). However, these estimates are based on the assumption that the bulk planetary $^{143}Nd/^{144}Nd$ ratio is identical to that of a chondritic uniform reservoir (CHUR) based on both C-chondrites and O-chondrites (*13*, *14*) and could be inaccurate if $\varepsilon^{143}_{Nd} \neq 0$ (Fig. 4).

The marked absence of measurable variations in r- and s-process isotopes for Sm and Nd (Figs. 2 and 3) justifies the use of both groups of chondrites in obtaining the CHUR $^{143}Nd/^{144}Nd$ ratio. However, unlike the $^{142}Nd/^{144}Nd$ ratio, the $^{143}Nd/^{144}Nd$ ratio is strongly dependent on the estimated CHUR Sm/Nd ratio. High-precision data for large (>0.2 g) bulk samples of chondrites [(*1*, *13*, *15*); table S6] indicate that the $^{143}Nd/^{144}Nd$ ratio of CHUR is not uniform at the ε level; it ranges from –2.1 to +0.8 ε units. There is no reason that CHUR should be best represented by an average of the measured chondrite data; it likely falls somewhere in the range given above. Our estimated range of CHUR (≡ bulk Earth) $^{142}Nd/^{144}Nd$, when taking $^{143}Nd/^{144}Nd$ variations of chondrites into account, is –18 ± 8 μ. This estimate is significantly different from the $^{142}Nd/^{144}Nd$ of the upper mantle and favors an early terrestrial differentiation and preservation of a low Sm/Nd reservoir (*1*).

It is evident that Earth, O-chondrites, EPB, and possibly the Moon have the same p-process contribution of ^{144}Sm and ^{146}Sm. In contrast, the C-chondrite parent bodies appear to have formed from nebular material with a somewhat lower contribution of p-nuclides. Because the C-chondrite parent bodies likely exist between 2.7 and 3.5 AU (outer asteroid belt) and the other objects come from regions between 1 and 2.4 AU, our data suggest that distant regions of the nebula received smaller amounts of the p-nuclides. This variation could be explained by the grain-size model (*16*), which invokes differential surface adsorption of p-nuclides on gas and/or dust particles whose size increases away from the core of the nebula. The yields of

all p-nuclides are approximately the same (*17*), and if their adsorption efficiencies on gas and/or dust are similar, C-chondrites should display a ~1 ε deficit in p-nuclides with respect to O-chondrites. For example, variations of this scale have been reported for ^{53}Mn; however, a radial gradient in ^{53}Mn between Earth and Vesta [e.g., (*18*)] is not supported by the Sm data. Recent studies suggest that ^{53}Mn could be produced by spallation reactions (*19*), in which case the Mn-Cr system would be decoupled from the Sm-Nd system. Data for p-process isotopes of Mo (^{92}Mo, ^{94}Mo) and Ru (^{96}Ru, ^{98}Ru) give apparently conflicting results (*20*–*23*); none of the studies finds the expected deficit in p-isotopes. This could be due to the influence of s- and r-process variations on the isotope pair used to correct for mass-dependent fractionation.

A recurrence interval (Δ) for SNe that provided the p-nuclides may be calculated assuming continuous uniform production of the nuclides with the last event occurring Δ years before solar system formation (SSF) (*24*):

$$\left(\frac{^{146}\mathrm{Sm}}{^{144}\mathrm{Sm}}\right)_{\mathrm{SSF}} = P_{^{146}\mathrm{Sm}/^{144}\mathrm{Sm}}\left(\frac{\bar{\tau}_{146}}{T_{\mathrm{UP}}}\right)\exp\left(\frac{-\Delta}{\bar{\tau}_{146}}\right)$$

where $P_{^{146}\mathrm{Sm}/^{144}\mathrm{Sm}}$ is the production ratio in a single event, T_{UP} is the period of uniform production, and $\bar{\tau}_{146}$ is the mean lifetime of ^{146}Sm = 149 My. For $P_{^{146}\mathrm{Sm}/^{144}\mathrm{Sm}}$ = 0.7 to 2.0 (*25*, *26*) and T_{UP} = 10 Gy [≡ age of the Galaxy (*27*)], Δ = 40 to 200 My. Recent studies suggest that the r-nuclides of atomic mass $A > 130$ were possibly created in the neutrino-driven wind from a neutron star produced by the accretion-induced collapse of a white dwarf in a binary system (*28*, *29*). The recurrence time scale for such stellar sites appears to be ~20 My (*27*). Evidently the frequency of production of p-nuclides is somewhat less than that of the heavy r-nuclides. It is similar to the frequency of production of light r-nuclides with $A \leq 130$ (*28*). Because the production of p- and r-nuclides is decoupled, the stellar sites that produce only p-nuclides likely contributed material to the solar nebula (*30*). Future models of planetary evolution must now consider the isotopic heterogeneity within the solar nebula.

References and Notes

1. M. Boyet, R. W. Carlson, *Science* **309**, 576 (2005); published online 16 June 2005 (10.1126/science.1113634).
2. K. Rankenburg, A. D. Brandon, C. R. Neal, *Science* **312**, 1369 (2006).
3. E. Zinner, S. Amari, R. S. Lewis, *Astrophys. J.* **382**, L47 (1991).
4. G. Wallerstein *et al.*, *Rev. Mod. Phys.* **69**, 995 (1997).
5. See supporting material on *Science* Online.
6. H. Hidaka, M. Ebihara, S. Yoneda, *Earth Planet. Sci. Lett.* **173**, 41 (1999).
7. G. P. Russ, D. S. Burnett, R. E. Lingenfelter, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **13**, 53 (1971).
8. H. Hidaka, M. Ebihara, S. Yoneda, *Earth Planet. Sci. Lett.* **180**, 29 (2000).
9. B. W. Stewart, D. A. Papanastassiou, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **143**, 1 (1996).
10. C. Arlandini *et al.*, *Astrophys. J.* **525**, 886 (1999).
11. K. Wisshak, F. Voss, F. Kappeler, L. Kazakov, G. Reffo, *Phys. Rev. C* **57**, 391 (1998).
12. M. C. Ranen, S. B. Jacobsen, *Lunar Planet. Sci. Conf.* **XXXVII**, 1832 (2006).
13. S. B. Jacobsen, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **50**, 139 (1980).
14. S. B. Jacobsen, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **67**, 137 (1984).
15. P. J. Patchett, J. D. Vervoort, U. Soderlund, V. J. M. Salters, *Earth Planet. Sci. Lett.* **222**, 29 (2004).
16. B. S. Meyer, D. D. Clayton, *Space Sci. Rev.* **92**, 133 (2000).
17. T. Hayakawa *et al.*, *Nucl. Phys. A* **758**, 525C (2005).
18. A. Shukolyukov, G. W. Lugmair, *Space Sci. Rev.* **92**, 225 (2000).
19. M. Gounelle *et al.*, *Astrophys. J.* **640**, 1163 (2006).
20. N. Dauphas, B. Marty, L. Reisberg, *Astrophys. J.* **565**, 640 (2002).
21. Q. Z. Yin, S. B. Jacobsen, K. Yamashita, *Nature* **415**, 881 (2002).
22. H. Becker, R. J. Walker, *Nature* **425**, 152 (2003).
23. J. H. Chen, D. A. Papanastassiou, G. J. Wasserburg, H. H. Ngo, *Lunar Planet. Sci. Conf.* **XXXV**, 1431 (2004).
24. D. N. Schramm, G. J. Wasserburg, *Astrophys. J.* **162**, 57 (1970).
25. E. Somorjai *et al.*, *Astron. Astrophys.* **333**, 1112 (1998).
26. A. Arnould, S. Goriely, *Phys. Rep.* **384**, 1 (2003).
27. Y. Z. Qian, G. J. Wasserburg, *Phys. Rep.* **333–334**, 77 (2000).
28. G. J. Wasserburg, M. Busso, R. Gallino, *Astrophys. J.* **466**, L109 (1996).
29. Y. Z. Qian, G. J. Wasserburg, *Astrophys. J.* **588**, 1099 (2003).
30. D. L. Lambert, *Astron. Astrophys. Rev.* **3**, 201 (1992).

31. This study has been made possible through generous sharing of samples and standards by G. J. Wasserburg. Funding was provided by NSF grants EAR-0130631 and EAR-0336405, a NASA NH Space Grant, and Dartmouth College. We thank C. Chen and J. D. Donoghue for providing technical development and support. Thorough reviews by A. Brandon, R. W. Carlson, G. J. Wasserburg, and an anonymous reviewer substantially improved the paper.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1131708/DC1
Materials and Methods
Fig. S1
Tables S1 to S6
References

23 June 2006; accepted 19 September 2006
Published online 5 October 2006;
10.1126/science.1131708
Include this information when citing this paper.

Barium Isotopes in Chondritic Meteorites: Implications for Planetary Reservoir Models

Michael C. Ranen* and Stein B. Jacobsen

High-precision barium isotope measurements yielded differences of up to 25 parts per million in the ^{137}Ba/^{136}Ba ratio and 60 parts per million in the ^{138}Ba/^{136}Ba ratio between chondrites and Earth. These differences probably arose from incomplete mixing of nucleosynthetic material in the solar nebula. Chondritic meteorites have a slight excess of supernova-derived material as compared to Earth, demonstrating that the solar nebula was not perfectly homogenized upon formation.

The model of (*1*), which proposed that our solar system formed from a hot, well-mixed, gaseous disk, has now been replaced with models that allow for the preservation of presolar materials [compare with (*2*)]. Although it is still generally thought that there would be no measurable variations in isotopic composition between bulk planetary bodies, it has been known since the mid-1970s that there are oxygen isotope differences between meteorite parent bodies Earth and Mars (*3*, *4*), proving that distinct isotopic reservoirs do exist between planetary bodies.

All the chemical elements heavier than Fe were formed from processes other than nuclear burning within stellar interiors. The three most important processes are the r, s, and p processes, though there may be multiple pathways for each of these that lead to even more varied compositions. A high–neutron-density environment is formed after a supernova explosion, giving rise to a rapid neutron addition process (r process). In an asymptotic giant branch star, the neutron density is not as high, and the slow neutron addition process (s process) is dominant (*5*). The p process is a small contribution and is probably formed by photodisintegration reactions associated with supernovae. Other processes, such as a neutron burst (n process) thought to be in the He shell of a supernova, have been suggested, based on the isotopic composition of special presolar grains such as silicon carbide (SiC) X grains (*6*, *7*). The n process produces a different isotopic composition as compared to the canonical r process (*8*).

Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: ranen@fas.harvard.edu

Clear evidence for the preservation of presolar isotopic signatures has been found in presolar grains, such as SiC and diamonds, representing mainly s-process materials, in chondritic meteorites (*9*, *10*). These grains have extreme isotopic variations as compared to the bulk meteorites. Refractory inclusions in some carbonaceous chondrites contain heavy-element isotopic anomalies, representing r-process excesses (*11*). If the solar nebula was hot and well-mixed, these grains and their isotopic effects would have been completely destroyed. The existence of these grains implies that either the solar nebula was cooler than thought or these grains were added later.

It has recently been confirmed that there are small variations in the $^{142}Nd/^{144}Nd$ ratio between Earth and bulk chondritic meteorites (*12*, *13*). Specifically, it was found that the $^{142}Nd/^{144}Nd$ ratio is anywhere from 10 to 40 ppm lower in chondrites than in the bulk silicate Earth. This difference was attributed to the decay of now extinct ^{146}Sm [with a half-life ($T_{1/2}$) of 103 million years (My)] in reservoirs with distinct Sm/Nd ratios (*13*). The model in (*13*) calls for the formation and sequestration of an enriched liquid (having a low Sm/Nd ratio) at the end of magma ocean crystallization on Earth after 30 My. The Moon appears to have a value of –20 ppm (*14*), similar to that of the ordinary chondrites (*13*), whereas Earth-like values reported by another group (*15*) complicate the issue further.

The conclusion of (*13*) is based on the assumption that the Earth and chondritic meteorite parent bodies had the same initial $^{142}Nd/^{144}Nd$ ratio. There are at least two other ways to produce differences in the $^{142}Nd/^{144}Nd$ ratio between these bodies. Incomplete mixing in the solar nebula would cause the $^{142}Nd/^{144}Nd$ ratio to have the largest anomaly of all Nd isotopes, because ^{142}Nd is an abundant nuclide made only by the s process. Variations in the $^{146}Sm/^{144}Sm$ ratio between planetary bodies may exist, affecting the $^{146}Sm/^{144}Nd$ ratio of a bulk planet and subsequently the amount of ^{142}Nd produced by decay. Thus, it is imperative to measure the isotopic composition of other heavy elements having contributions from both the s and r processes but not having a substantial contribution from radioactive decay (*16*).

Ba is one such element, sufficiently abundant in chondrites (~2 to 4 ppm) to be measured with high precision. It has seven isotopes: two are p-process only (130 and 132), two are s-process only (134 and 136), and three are formed by both s and r processes (135, 137, and 138). ^{135}Ba may also be produced by decay from now extinct ^{135}Cs ($T_{1/2}$ = 2.3 My) (*17*, *18*), but the initial abundance of ^{135}Cs in the early solar system is uncertain [compare with (*19*)]. Both r and s process–derived Ba nuclear anomalies exist in components of carbonaceous chondrites. Refractory inclusions in the Allende meteorite contain evidence for positive excesses in r-process isotopes of Ba and Nd (*11*). Mainstream SiC grains in acid residues from the Murchison meteorite show large s-process–derived Ba isotope signatures (*10*). Other unusual SiC grains (X grains) show a signature enriched in ^{137}Ba and more enriched in ^{138}Ba, which probably formed from a neutron burst effect in a supernova (*7*).

We measured the Ba isotopic compositions of six chondritic meteorites [Allende (CV3), Murchison (CM2), Grady (H3.7), Guarena (H6), Peace River (L6), and Bruderheim (L6)] with a next-generation multicollector thermal ionization mass spectrometer (a GV Instruments Isoprobe-T) and compared the results to Ba from measurements of Earth (*20*). We also measured an acetic acid leach of the Murchison meteorite corresponding to 50% of the Ba in the bulk sample (Ba concentration of 2.21 ppm as compared to 4.46 ppm in our whole-rock sample). To represent Ba from Earth, we used ultrapure Ba from a 1000-ppm standard solution from SPEX (H-Ba#1). We also measured Ba separated from a terrestrial andesite, the U.S. Geological Survey standard AGV-1. AGV-1 was subjected to the same chemical separation procedures used for the meteorites to investigate whether there were unusual effects resulting from our purification procedures. The data were normalized to the $^{134}Ba/^{136}Ba$ ratio, the components of which are both s process–only nuclides, in order to investigate possible variations in the nuclides having contributions from both the r and s process; namely ^{135}Ba, ^{137}Ba, and ^{138}Ba. Our best values for H-Ba#1 are $^{134}Ba/^{136}Ba$ = 0.30781144, $^{135}Ba/^{136}Ba$ = 0.83934114, $^{137}Ba/^{136}Ba$ = 1.4290088, and $^{138}Ba/^{136}Ba$ = 9.12875 (*20*). All meteorites have close-to-normal levels of ^{135}Ba but small excesses in ^{137}Ba and larger excesses in ^{138}Ba up to 60 ppm (Fig. 1). In contrast, AGV-1 is within the error of the terrestrial values defined by the H-Ba#1 standard.

Our results show that similar excesses exist in ^{137}Ba and ^{138}Ba for carbonaceous chondrites (Allende and Murchison) and the primitive ordinary chondrite (Grady) as compared to Earth. Class 6 ordinary chondrites (Bruderheim, Peace River, and Guarena) exhibit smaller excesses in ^{138}Ba as compared to Earth, and ^{137}Ba effects are hard to resolve. Whereas Murchison and Allende both have significant but different levels of hard-to-dissolve presolar grains such as SiC, graphite, and diamonds, Grady has essentially none left (*21*). The class 6 ordinary chondrites have all been subject to the highest thermal metamorphism in meteorite parent bodies, resulting in complete equilibration of minerals and no preservation of presolar grains. Thus, it is unlikely that the Ba isotope differences that we observe between Earth and chondrites are due to incomplete dissolution of presolar grains in the samples studied. Also, isotope fractionation processes on Earth or the chondrite parent bodies will not cause these variations. The most likely explanation is that this difference reflects incomplete mixing in the solar nebula of diverse

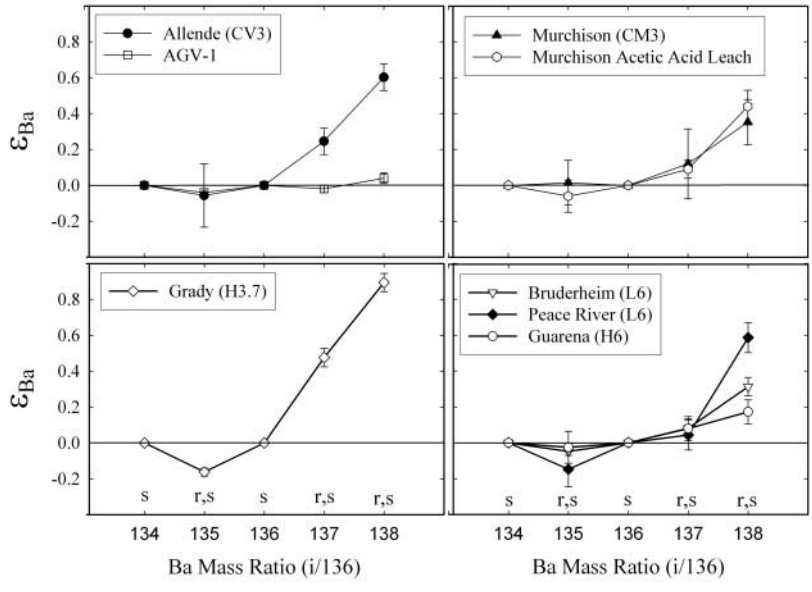


Fig. 1. All meteorite data were normalized to our terrestrial standard. AGV-1, a terrestrial andesite, is shown for comparison. ε_{Ba} is defined as the deviation in Ba isotope abundance in parts per 10,000 from a terrestrial standard after the raw data have been normalized to $^{134}Ba/^{136}Ba$, $\left(\frac{\left(\frac{^{13x}Ba}{^{136}Ba}\right)_{sample}}{\left(\frac{^{13x}Ba}{^{136}Ba}\right)_{standard}} - 1\right) \cdot 10{,}000$, where 13x represents the selected isotope. Error bars are 2 SDs based on the external reproducibility. s and r indicate whether each isotope is made by the s or r process; i/136 represents a particular Ba isotope relative to ^{136}Ba. See the supporting online material for more details about the measurements.

nucleosynthetic components. Although a 1-g sample is not necessarily representative of the average of the parent bodies for individual meteorite classes, it is clear that the parent bodies of all chondritic meteorite classes that we investigated exhibit Ba isotopic compositions that are distinctly different from Earth's.

The similarity of the bulk meteorite Ba isotope pattern to those of Murchison X grains (*7*), interpreted as being formed by a neutron burst (n process) in a star undergoing a supernova explosion, suggests that chondritic meteorites are enhanced in such a component as compared to Earth. The SiC X grains from Murchison that have a similar Ba isotopic pattern (Fig. 2) (*7*) are too rare in this meteorite to cause measurable differences in bulk meteorite samples (*22*). Previous results on Mo isotopes (*23*) also showed excesses of supernova components as compared to Earth and, in particular, one of these resembled an n-process component for Mo. Results on Zr isotopes (*24*) in carbonaceous chondrites show a large excess in ^{96}Zr, which is consistent with an n-process component (*7*). This makes it likely that the chondritic meteorite parent bodies accreted a small (<0.1%) excess of neutron burst material [with the use of Murchison X grains 113-2 and 113-3 (*7*) as a proxy for the n process] as compared to Earth (*25*). Thus, variations in the bulk isotopic composition of undifferentiated meteorites demonstrate that the material aggregating to make Earth and chondritic parent bodies was not completely homogenized nor was it all processed at temperatures high enough to erase the signatures of diverse presolar stellar sources.

^{138}Ba would be expected to anticorrelate with variations in ^{142}Nd for incomplete mixing of r- and n-process components (Fig. 3). Both carbonaceous and ordinary chondrites showed similar effects. The r- or n-process excesses seen in Ba from chondritic meteorites predict ^{142}Nd deficits in chondritic meteorites as compared to Earth (*26*). Because multiple supernova component excesses appear to be present in chondritic meteorites, there is not a straightforward approach for how to use measurements of isotopes for one element such as Ba to quantitatively predict what the initial ^{142}Nd/^{144}Nd isotopic composition of Earth was. Yet the overall anticorrelation makes it likely that the initial ^{142}Nd/^{144}Nd isotopic composition is one that would evolve to a value close to that observed in the modern silicate Earth, rather than to the 20-ppm–lower value favored by (*13*). However, the most recent estimate of the ^{142}Nd/^{144}Nd composition of the Moon as chondritic (*14*), which suggests that the Moon formed from Earth before it underwent early differentiation at 30 My as proposed by (*13*), complicates this further. Lunar samples exhibit variable ^{142}Nd/^{144}Nd ratios because of ^{146}Sm decay coupled with early lunar Sm/Nd fractionation and neutron capture effects that are difficult to correct for at the required level of precision. Therefore, it is more difficult to arrive at a reliable bulk lunar ^{142}Nd/^{144}Nd ratio.

Magma oceans almost certainly existed on the Moon and Mars and most likely existed on Earth as well. Although it is likely that early differentiation of a terrestrial magma ocean produced ^{142}Nd/^{144}Nd isotopic variations (*26*, *27*), it is also likely that, after ^{146}Sm became extinct (~500 My after the formation of the solar system), convection mixed the mantle well enough to destroy most of the heterogeneities caused by fractionation in a magma ocean (*28*). However, isotopic heterogeneities between planetary bodies caused by incomplete mixing of the solar nebula do exist; thus, one must be careful in predicting initial isotope ratios for Earth based on those in chondritic meteorites.

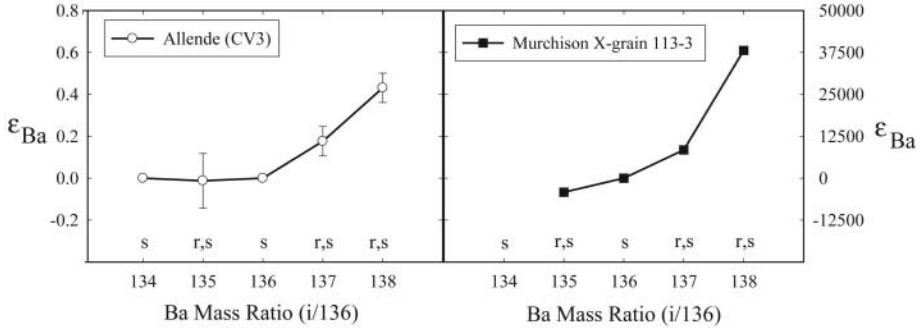


Fig. 2. Comparisons between Ba isotope deviations in parts per 10,000 for Allende with those of the Murchison X grain 113-3 [data from (*7*)]. There is a marked similarity in the pattern of the two (though the X grain is at a much larger scale). The relative enrichment in the X-grain data matches our bulk meteorite data much more strongly than what is thought to be a typical r-process component (*11*). Error bars indicate 2 SDs based on external reproducibility.

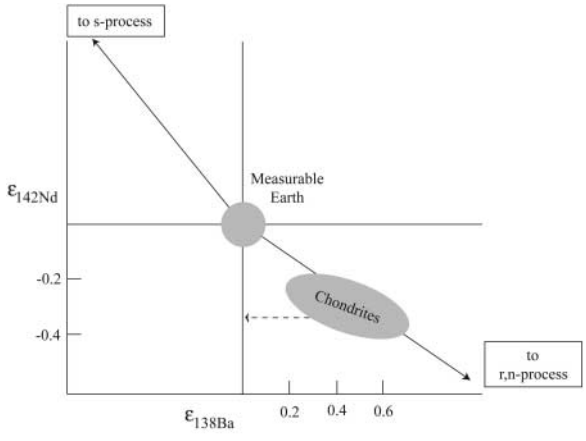


Fig. 3. Cartoon showing anticorrelation between negative deviation in ^{142}Nd and positive deviation in ^{138}Ba between the measurable Earth and chondritic meteorites. If early fractionation on Earth and the subsequent decay of ^{146}Sm were the only causes for ^{142}Nd differences, then all Ba data would plot along a vertical line going through $\varepsilon_{138Ba} = 0$, as indicated by the dashed arrow. The ^{138}Ba data roughly correlate with the ^{142}Nd data pointing toward a supernova (r-process or neutron burst) excess in meteoritic material.

References and Notes

1. A. G. W. Cameron, *Icarus* **1**, 339 (1963).
2. P. Cassen, K. M. Chick, in *Astrophysical Implications of the Laboratory Study of Presolar Materials*, T. J. Bernatowicz, E. K. Zinner, Eds. (American Institute of Physics Conference Proceedings 402, Woodbury, NY, 1997), pp. 697–719.
3. R. N. Clayton, N. Onuma, T. K. Mayeda, *Earth Planet. Sci. Lett.* **30**, 10 (1976).
4. R. N. Clayton, in *Treatise on Geochemistry*, H. D. Holland, K. K. Turekian, Eds. (Elsevier, Amsterdam, 2003), vol. 1, chap. 6.
5. E. M. Burbidge, G. R. Burbidge, W. A. Fowler, F. Hoyle, *Rev. Mod. Phys.* **29**, 547 (1957).
6. B. S. Meyer, D. D. Clayton, L.-S. The, *Astrophys. J.* **540**, L49 (2000).
7. M. J. Pellin *et al.*, *Lunar Planet. Sci.* **XXXI**, A1917 (2000).
8. W. M. Howard, B. S. Meyer, D. D. Clayton, *Meteoritics* **27**, 404 (1992).
9. U. Ott, F. Begemman, J. M. Yang, S. Epstein, *Nature* **332**, 700 (1988).
10. E. Zinner, S. Amari, R. S. Lewis, *Astrophys. J.* **382**, L47 (1991).
11. M. T. McCulloch, G. J. Wasserburg, *Astrophys. J.* **220**, L15 (1978).
12. C. L. Harper Jr., S. B. Jacobsen, *Lunar Planet. Sci.* **XXIV**, A607 (1993).
13. M. Boyet, R. Carlson, *Science* **309**, 576 (2005).
14. K. Rankenburg, A. D. Brandon, C. R. Neal, *Science* **312**, 1369 (2006).
15. L. E. Borg, M. Wadhwa, *Lunar Planet. Sci.* **XXXVII**, A1154 (2006).
16. Nd isotopes are not ideal to study bulk isotope variations in meteorites. ^{142}Nd and ^{143}Nd are both influenced by the radioactive decay of Sm isotopes, ^{144}Nd and ^{146}Nd are used for normalization, and ^{148}Nd and ^{150}Nd are small and further away from the normalizing ratios. This leaves only ^{145}Nd, which is a small isotope of Nd.
17. H. Hidaka, Y. Ohta, S. Yoneda, J. DeLaeter, *Earth Planet. Sci. Lett.* **193**, 459 (2001).
18. H. Hidaka, Y. Ohta, S. Yoneda, *Earth Planet. Sci. Lett.* **214**, 455 (2003).
19. S. B. Jacobsen, M. C. Ranen, *Lunar Planet. Sci.* **XXXVII**, A2241 (2006).

20. See the supporting material on *Science* Online for specific information about the data.
21. G. R. Huss, in *Astrophysical Implications of the Laboratory Study of Presolar Materials*, T. J. Bernatowicz, E. K. Zinner, Eds. (American Institute of Physics Conference Proceedings 402, Woodbury, NY, 1997), pp. 721–746.
22. SiC X grains make up at most 45 parts per billion of Murchison (*18*). The typical Murchison X grain has an enrichment of 38,000ε in ^{138}Ba (*7*) and has a Ba concentration of approximately twice that of CI Ba (9 ppm) (*8*). This will contribute only 0.3 ppm excess in the whole rock of Murchison.
23. Q. Yin, S. B. Jacobsen, K. Yamashita, *Nature* **415**, 881 (2002).
24. Q. Yin, S. B. Jacobsen, J. Blichert-Toft, P. Telouk, F. Albarede, *Lunar Planet. Sci.* **XXXII**, A2128 (2001).
25. Excesses in these X grains as compared to Earth are 38,080ε for ^{138}Ba/^{136}Ba and 22,230ε for ^{96}Zr/^{94}Zr in grain 113-3 and 3310ε for ^{95}Mo/^{96}Mo in grain 113-2 (*7*). Adding between 0.1 and 0.001% of these components to an isotopically normal sample can give the observed variations seen in bulk Murchison (0.35ε in ^{138}Ba/^{136}Ba, 48ε in ^{96}Zr/^{94}Zr, and 1.80ε in ^{95}Mo/^{96}Mo) (*20*, *23*).
26. C. L. Harper Jr., S. B. Jacobsen, *Nature* **360**, 728 (1992).
27. G. Caro, B. Bourdon, J. L. Birck, S. Moorbath, *Geochim. Cosmochim. Acta* **70**, 164 (2006).
28. S. B. Jacobsen, C. L. Harper Jr., *Geochim. Cosmochim. Acta* **60**, 3747 (1996).
29. This work was done under a NASA Origins of Solar System Grant (NAG5-13164) and a NASA Cosmochemistry Grant (NNG04GG06G) to S.B.J. The comments of two anonymous reviewers greatly improved this manuscript.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1132595/DC1
Materials and Methods
Fig. S1
Tables S1 to S9
References

17 July 2006; accepted 15 September 2006
Published online 5 October 2006;
10.1126/science.1132595
Include this information when citing this paper.

From Plant Traits to Plant Communities: A Statistical Mechanistic Approach to Biodiversity

Bill Shipley,[1]* Denis Vile,[1,2] Éric Garnier[2]

We developed a quantitative method, analogous to those used in statistical mechanics, to predict how biodiversity will vary across environments, which plant species from a species pool will be found in which relative abundances in a given environment, and which plant traits determine community assembly. This provides a scaling from plant traits to ecological communities while bypassing the complications of population dynamics. Our method treats community development as a sorting process involving species that are ecologically equivalent except with respect to particular functional traits, which leads to a constrained random assembly of species; the relative abundance of each species adheres to a general exponential distribution as a function of its traits. Using data for eight functional traits of 30 herbaceous species and community-aggregated values of these traits in 12 sites along a 42-year chronosequence of secondary succession, we predicted 94% of the variance in the relative abundances.

Ecological models of community assembly have traditionally been conceived through the lens of population ecology (*1*–*4*). A classic example is the Lotka-Volterra system of equations in which the per capita growth rate of each of S species is linked through an S-by-S "community" matrix of interaction coefficients. This quickly becomes impractical. Interaction coefficients are difficult to measure and the number of binary interactions, even if unaffected by environmental variation, increases with the square of the number of species in the regional pool. Although such models can qualitatively describe community assembly in very simple empirical systems involving a few unicellular species in well-mixed and simple environments (*3*, *5*, *6*), they have not been able to do this in more complex field situations involving many species. Empirical attempts to predict relative abundance from plant traits without any theoretical context have largely failed (*7*). A more recent demographic approach is to assume that the per capita growth rates or average individual fitness of all species in a metapopulation are equal, leading to neutral models of community structure (*4*, *8*). Neutral models cannot predict which species will have which relative abundance; indeed, they imply that this is impossible because the dynamics are completely random. Furthermore, their community-level patterns (*9*) and the assumption of average individual fitness of all species in the metacommunity (*10*) have been rejected by empirical data.

We adopted the view of community assembly as a process of species sorting, leading to the concept of community assembly by environmental filters (*11*–*13*). A pool of S species determines those that can disperse to a site. A sorting of these species by the environment, including the environment as modified by the interacting species, determines the relative abundances of those that actually occur at a site. This is simply a result of the nonrandom demographic processes of dispersal, growth, survival, and reproduction of individuals within and between species, as determined by the functional traits they possess, when the interspecific variations of such functional traits are greater than their intraspecific variations.

Species sorting by the environment is a process of selection over short (ecological) time scales, resulting in changes in species composition over an environmental gradient. The breeder's equation of quantitative genetics (*14*) (Eq. 1b or its multivariate equivalent) describes how the mean value of a quantitative trait j of genotype i, occurring with proportion p_{ik} at site k in a population of n genotypes at time x, changes over the time period during which a selection event occurs. The amount of change in the mean trait value ($\overline{t_j}$) after the selection event is a function of its heritability (h^2, the slope of a regression of offspring trait values on midparent trait values) and the force of selection [$S_j(x)$] for that trait at time x:

$$\overline{t_j(x)} = \sum_{i=1}^{n} p_{ik}(x) t_{ij} \qquad (1a)$$

$$\left[\overline{t_j(1)} - \overline{t_j(0)}\right] = h^2(0) S_j(0) \qquad (1b)$$

Equation 1 is usually applied to selection between genotypes of a single species, but it also applies even if the genotypes occur in different species. Because intraspecific variances of our functional traits are much smaller than their interspecific variances (*15*), we assumed that such trait values can be approximated as species-specific attributes. We also assumed that intraspecific genotypic evolution of the trait values during ecological time is insignificant relative to preexisting interspecific variation. Given these two assumptions, selection will primarily occur between species, and the heritability will be close to unity, although this is not a requirement of the model. Iterating over time gives

$$\overline{t_j(x)} = \sum_{k=0}^{x-1} h^2(k) S_j(k) + \overline{t_j(0)} \qquad (2)$$

The relative abundance of each species in the regional pool, and therefore the value of $\overline{t_j(0)}$,

[1]Département de Biologie, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada. [2]Centre d'Écologie Fonctionnelle et Évolutive, CNRS, UMR 5175, 1919 Route de Mende, 34293 Montpellier Cedex 5, France.

*To whom correspondence should be addressed. E-mail: bill.shipley@usherbrooke.ca

will be the same for all sites in the region. Once community assembly begins, any differences between sites in $\overline{t_j(x)}$—called "community-aggregated" traits (*16*)—will reflect different selective pressures over time or across environmental gradients. These community-aggregated traits reflect the constraints on community assembly imposed by the process of species sorting.

Once plants are at a site, the biomass (abundance) of each will be proportional to the total amount of resource units that each species is able to capture at that point in time. The total number of such captured resources $N_k(x)$ at site k at time x, and the number that are captured by each species $n_{ik}(x)$, defines the abundance structure of the community: $\{n_{1k}(x), n_{2k}(x), \ldots, n_{Sk}(x)\}$. Defining relative abundance as $p_{ik}(x) = n_{ik}(x)/N_k(x)$ describes community structure independent of site productivity.

Assume first a strict equivalence of species such that each species is equally likely to disperse and capture resources. There would be no species sorting and community assembly would be purely random. The number of different ways (W) that the N resource units, comprising the living biomass at a site, could be partitioned into a particular community structure is

$$W = \frac{N!}{n_1!n_2!\ldots n_S!} \quad (3)$$

Site and time indices are suppressed for simplicity. N will be less than the number of atoms of carbon and mineral elements in the biomass because these must respect the stoichiometry of the tissues, but it will be very large. Taking logarithms, Sterling's approximation $[\log(n!) \xrightarrow[n\to\infty]{} n\log(n)-n]$ can be rearranged to form the following equations:

$$\ln(W) = N\ln(N) - \sum_{i=1}^{S} [n_i \ln(n_i)] \quad (4a)$$

$$\frac{\ln(W)}{N} = -\sum_{i=1}^{S} p_i \ln(p_i) \quad (4b)$$

The most likely community composition is the one maximizing W and, given large N, this is equivalent to maximizing Shannon's (*17*) index of information entropy. If all species are not strictly equivalent, because their traits allow individuals of some species to capture more resources, then not all partitions will be equally likely to occur. The species sorting, based on functional traits, results in particular community-aggregated trait values, and therefore excludes partitions that contradict such values. A maximally random community (**p**) constrained by these community-aggregated values (Eqs. 1 and 2) are the values of p_i that maximize W conditional on all $\overline{t_j}$ plus normalization ($\sum_{i=1}^{S} p_i = 1$). This is found, using the method of Lagrangian multipliers, by defining a new system of equations: $f(p_i) = \ln(W) - \lambda_o(1 - \sum p_i) - \sum_{j=1}^{T} \lambda_j \left(\overline{t_j} - \sum_{i=1}^{S} p_i t_{ij} \right)$ and solving for the T+1values of λ_j when the partial derivates are set to zero. The general solution to this problem (*18*), from the maximum entropy formalism, is Eq. 5, where $\hat{p}_i$ is the predicted relative abundance of species i, possessing the T functional traits $\mathbf{t_i} = \{t_{i1},\ldots,t_{iT}\}$. The maximum likelihood values of the nine estimated values of λ in our study were obtained using the Improved Iterative Scaling algorithm of Della Pietra *et al.* (*19*, *20*).

$$\hat{p}_i = \frac{e^{\left(\lambda_o - \sum_{j=1}^{T} \lambda_j t_{ij}\right)}}{\sum_{i=1}^{S} \left[e^{\left(\lambda_o - \sum_{j=1}^{T} \lambda_j t_{ij}\right)} \right]} \quad (5)$$

We conducted our study in 12 vineyards that had been abandoned between 2 and 42 years previously, within a 4-km^2 area of southern France. The aboveground dry biomass of all species in each of four plots measuring 0.25 by 0.25 m was used to estimate relative abundances for each site. Details of the sites and the methods of vegetation sampling are given in (*16*). We measured eight functional traits on 30 species, representing at least 80% of the total biomass

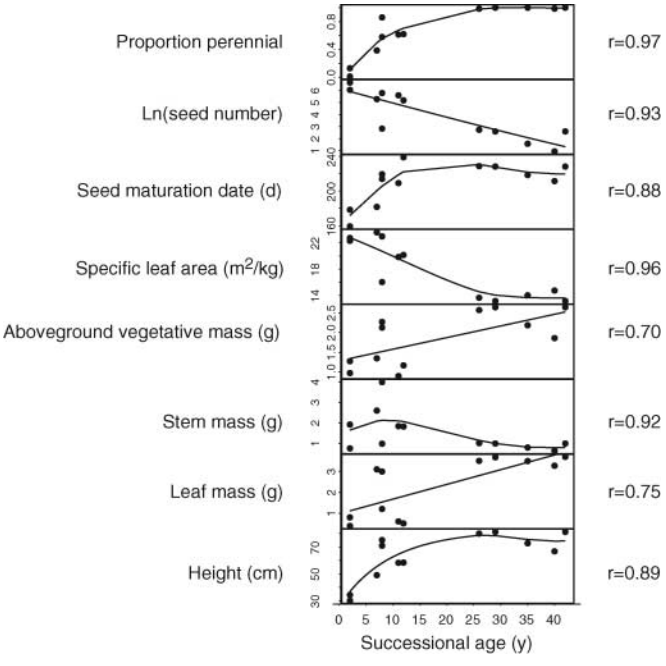


Fig. 1. Each point represents one of 12 sites in a 42-year chronosequence of secondary succession in southern France after vineyard abandonment. The dependent variables are the community-aggregated trait values of the vegetation at that site: proportional perenniality (0 = all biomass belongs to annual species, 1 = all biomass belongs to perennial species), Ln-transformed number of seeds per plant, julian days until seed maturation, specific leaf area (m^2/kg), aboveground prereproductive vegetative dry mass (g), stem dry mass (g), leaf dry mass (g), postreproductive height (cm). Lines show the predicted values from a cubic-spline regression and the Pearson correlation coefficient (r) relates observed and predicted values.

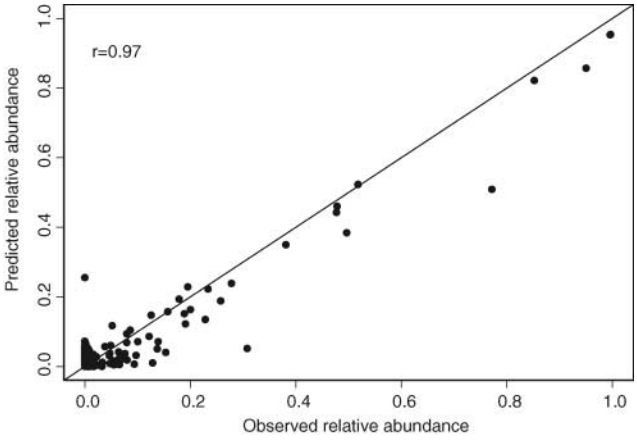


Fig. 2. Observed and predicted relative abundances of 30 species in 12 sites during a 42-year chronosequence of secondary succession. Predicted values are those obtained by maximizing the Shannon entropy conditional on the measured community-aggregated trait values of each site, shown in Fig. 1.

over the chronosequence, and calculated the community-aggregated traits as defined in Eq. 1a. Details of the measurements of these traits and their values are given in (*21*). Figure 1 shows the eight community-aggregated traits of these 12 sites during the chronosequence. Using these observed community-aggregated traits, the maximum entropy estimates predicted 94% of the observed variation in the actual relative abundances of these 30 species in the 12 sites (Fig. 2).

For generality, it is preferable to predict the community-aggregated traits, given information on environmental conditions, rather than measuring them. Therefore, we predicted the community-aggregated traits in each site from its successional age using cubic-spline regression (Fig. 1). Using these predicted values, we obtained the predicted community structure for each age (Fig. 3, B and D). The observed successional dynamics of each species, smoothed using a cubic-spline smoother, are shown in Fig. 3, A and C; we also show this on a $\log_e$ scale (Fig. 3, C and D) to emphasize the rare species. The observed and predicted successional dynamics are quite close ($r = 0.96$).

Curiously, given the historical dominance of the demographic Lotka-Volterra equations, Volterra recognized the difficulties of this approach and even considered a statistical mechanistic approach (*22*). Very few authors have followed his lead (*23*–*31*). Classical statistical mechanics ignores, but does not deny, the detailed dynamics of atomic collisions and assumes a random allocation of atoms to energy states whose total (and therefore mean) energy is constrained. Our model ignores, but does not deny, the details of resource allocation, population dynamics, stochastic processes, and species interactions. It assumes a constrained random allocation of resource units to species; the constraints are generated by natural selection and are quantified by the community-aggregated trait values. The relative abundance of each species in a species pool is therefore a function of how closely its functional traits agree with the community-aggregated traits.

Determining whether our model is generally successful will require further testing. If successful, it could provide a quantitative link between functional ecology, community ecology, and biogeography, while still being applicable in the field. Furthermore, it could predict which species could successfully invade new environments and how plant communities will change after changes in environments. Because it includes components of both niche-based and neutral models, it could also point to a way of reconciling these two different conceptions of community assembly. Realizing these potentials will require further empirical testing, a consistent quantification of major environmental gradients, and a demonstration of generality in the patterns of community-aggregated traits along such gradients.

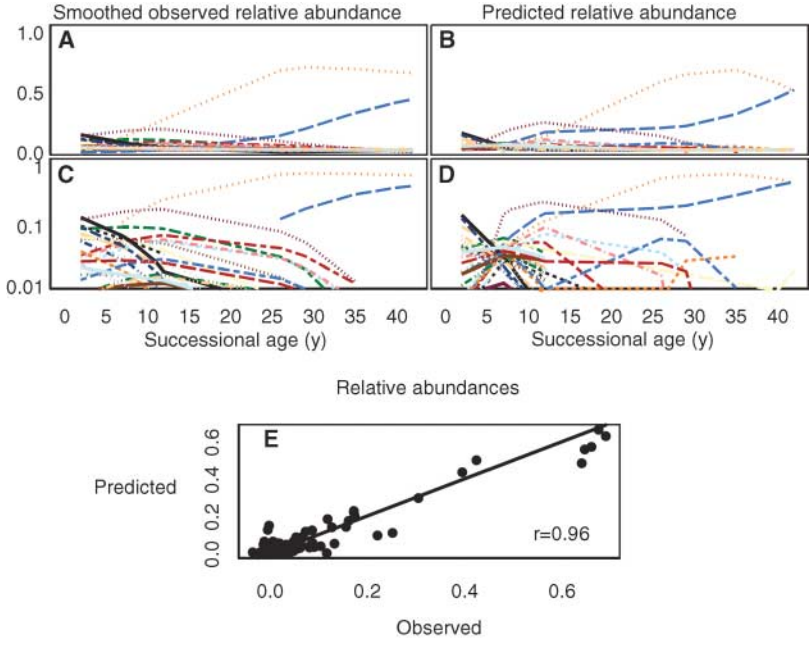


Fig. 3. (**A** and **C**) Observed relative abundance of each of 30 species during a 42-year chronosequence of secondary succession. Curves for each species were obtained using a cubic-spline regression. (**B** and **D**) Predicted relative abundance of each species, obtained by maximizing the Shannon entropy conditional on the measured community-aggregated trait values of each site, shown in Fig. 1. (A) and (B) show relative abundances on an arithmetic scale and (C) and (D) show the same data on a logarithmic scale in order to emphasize the rarer species. (**E**) The smoothed observed relative abundances of each species at each successional age plotted against the relative abundances predicted by maximizing the Shannon entropy conditional on the measured community-aggregated trait values at each successional age.

References and Notes

1. G. F. Gause, *The Struggle for Existence* (Williams and Wilkins, Baltimore, 1934).
2. R. H. MacArthur, *Geographical Ecology: Patterns in the Distribution of Species* (Harper and Row, New York, 1972).
3. D. Tilman, *Resource Competition and Community Structure.* (Princeton Univ. Press, Princeton, NJ, 1982).
4. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
5. J. H. Vandermeer, *Ecology* **50**, 362 (1969).
6. J. Passarge, S. Hol, M. Escher, J. Huisman, *Ecol. Monogr.* **76**, 57 (2006).
7. B. R. Murray, P. H. Thrall, A. M. Gill, A. B. Nicotra, *Austral Ecol.* **27**, 291 (2002).
8. G. Bell, *Ecology* **86**, 1757 (2006).
9. B. J. McGill, B. A. Maurer, M. D. Weiser, *Ecology* **87**, 1411 (2006).
10. W. S. Harpole, D. Tilman, *Ecol. Lett.* **9**, 15 (2006).
11. P. A. Keddy, *J. Veg. Sci.* **3**, 157 (1992).
12. E. Weiher, P. A. Keddy, *Oikos* **73**, 323 (1995).
13. J. P. Grime, *Plant Strategies, Vegetation Processes, and Ecosystem Properties* (John Wiley & Sons, New York, ed. 2, 2001).
14. D. A. Roff, *Evolutionary Quantitative Genetics* (Chapman & Hall, New York, 1997).
15. E. Garnier *et al.*, *New Phytol.* **152**, 69 (2001).
16. E. Garnier *et al.*, *Ecology* **85**, 2630 (2004).
17. C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication* (Univ. of Illinois Press, Urbana, IL, 1949).
18. E. T. Jaynes, *Probability Theory: The Logic of Science.* (Cambridge Univ. Press, Cambridge, UK, 2003).
19. S. Della Pietra, V. Della Pietra, J. Lafferty, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 380 (1997).
20. A program to obtain our parameter estimates is available online (http://pages.usherbrooke.ca/jshipley/recherche/my%20programs.htm).
21. D. Vile, B. Shipley, E. Garnier, *Ecology* **87**, 504 (2006).
22. V. Volterra, *Leçons sur la Théorie Mathématique de la Lutte pour la Vie* (Gauthier-Villars, Paris, 1931).
23. M. Plank, *SIAM J. Appl. Math.* **59**, 1540 (1999).
24. M. Plank, *Nonlinearity* **9**, 887 (1996).
25. M. Plank, *J. Math. Phys.* **36**, 3520 (1995).
26. E. H. Kerner, *Gibbs Ensemble: Biological Ensemble* (Gordon and Brench, New York, 1964).
27. D. Lurie, J. Wasenburg, in *Thermodynamics and Regulation of Biological Processes*, I. Lambrecht, A. I. Zotin, Eds. (De Gruyter, Berlin, 1984), pp. 257–271.
28. A. P. Levich, in *Lectures in Theoretical Biology*, T. Tiivel, Ed. (Valgus, Tallinn, Estonia, 1988) pp. 121–127.
29. V. L. Alexeyev, A. P. Levich, *Bull. Math. Biol.* **59**, 649 (1997).
30. A. P. Levich, *Ecol. Modell.* **131**, 207 (2000).
31. B. A. Maurer, *Untangling Ecological Complexity: The Macroscopic Perspective* (Univ. of Chicago Press, Chicago, 1999).

32. This research was financially supported by the Natural Sciences and Engineering Research Council of Canada.

15 June 2006; accepted 31 August 2006
Published online 5 October 2006;
10.1126/science.1131344
Include this information when citing this paper.

ERRATUM

Post date 13 April 2007

Reports: "From plant traits to plant communities: a statistical mechanistic approach to biodiversity" by B. Shipley *et al.* (3 Nov. 2006, p. 812). In the denominator of Eq. 5 on page 813, the expression in the parentheses should have been set as an exponent of *e*. The correct equation is

$$\hat{p}_i = \frac{e^{\left(\lambda_0 - \sum_{j=1}^{T} \lambda_j t_{ij}\right)}}{\sum_{i=1}^{S} \left(e^{\left(\lambda_0 - \sum_{j=1}^{T} \lambda_j t_{ij}\right)} \right)}$$

FG-Rich Repeats of Nuclear Pore Proteins Form a Three-Dimensional Meshwork with Hydrogel-Like Properties

Steffen Frey,[1,2] Ralf P. Richter,[3,4] Dirk Görlich[1,2]*

Nuclear pore complexes permit rapid passage of cargoes bound to nuclear transport receptors, but otherwise suppress nucleocytoplasmic fluxes of inert macromolecules ≥30 kilodaltons. To explain this selectivity, a sieve structure of the permeability barrier has been proposed that is created through reversible cross-linking between Phe and Gly (FG)–rich nucleoporin repeats. According to this model, nuclear transport receptors overcome the size limit of the sieve and catalyze their own nuclear pore-passage by a competitive disruption of adjacent inter-repeat contacts, which transiently opens adjoining meshes. Here, we found that phenylalanine-mediated inter-repeat interactions indeed cross-link FG-repeat domains into elastic and reversible hydrogels. Furthermore, we obtained evidence that such hydrogel formation is required for viability in yeast.

Cell nuclei import all their proteins from the cytoplasm and, in return, supply the cytoplasm with nuclear products such as ribosomes, mRNAs, or transfer RNAs. Because interphase nuclei are enclosed by the nuclear envelope, all exchange between the compartments proceeds through the embedded nuclear pore complexes (NPCs). NPCs are equipped with a permeability barrier (*1*, *2*) against uncontrolled nucleocytoplasmic fluxes, which otherwise would have detrimental consequences, e.g., for the ordered course of gene expression (*1*). Objects larger than ~30 to 40 kD are restricted or even excluded from passing the barrier unless they are bound by cognate nuclear transport receptors (NTRs) (*1*, *3*, *4*). Because the actual NPC passage of cargo-NTR complexes is a reversible process, unidirectional cargo transport requires active cargo release from NTRs in the destination compartment.

Facilitated NPC passage requires interactions of NTRs with so-called FG repeats (*5*–*9*) that are found in numerous nuclear pore proteins (*10*), which suggests that the permeability barrier contains or is even built of these repeats [discussed in (*11*, *12*)]. Such repeat domains have up to 50 repeat units (*10*), appear intrinsically unfolded (*13*), and contain short clusters of hydrophobic amino acids [such as FSFG or GLFG in single-letter amino acid code (*14*)] separated by hydrophilic spacers.

Various models have been proposed to explain how these repeat domains create the permeability barrier. The "virtual gating" (*15*) and the "oily spaghetti" (*2*) models suggest that the FG-repeat domains repel molecules that should be excluded from passage. The "selective-phase" model (*11*, *12*) is based on the consideration that a simple binding of NTRs to such repeats would only delay their NPC passage. It assumes that the barrier is a sievelike structure, whereby meshes form through hydrophobic interactions between the FG-repeats' hydrophobic clusters, while the mesh size sets the size limit for passive exclusion. According to this model, NTRs can overcome this size restriction because their binding to hydrophobic clusters competes the noncovalent inter-repeat cross-links and, thereby, transiently opens adjacent meshes of the sieve (see fig. S1).

The models differ foremost in the question of whether the permeability barrier is tightened by inter-repeat interactions or not. If no interactions occurred, aqueous solutions of FG-repeat domains should behave like (viscous) fluids (Fig. 1, right inset). In contrast, if cross-linking prevented a free sliding between the linear polymers, elastic hydrogels should form (Fig. 1, left inset). Provided the cross-links remain intact, such a gel would respond to mechanical forces only by transient deformation and would readopt its original shape once the strain is relieved. We decided to test these predictions for the N-terminal "fsFG"-repeat domain of the essential yeast nucleoporin Nsp1p (*16*), which comprises 18 regular FSFG-repeats and 16 less regular FG-repeats (fig. S2).

The starting point for gelling a polymer into a homogeneous hydrogel is typically a homogeneous aqueous solution. Agarose or gelatin, for example, is first dissolved in hot water, and the respective gels are obtained after cooling. We found that, in our case, the solid-to-gel transition was best triggered by a pH shift. At high pH, the negative net charge of the Nsp1 fsFG-repeat domain is apparently repulsive enough to counteract gelling. Strikingly, however, when alkaline solutions with ≥ 8 mg/ml fsFG repeats were brought to physiological pH (where the net charge of the repeat units is close to zero), transparent and, hence, homogeneous gels formed (*17*). These gels were elastic enough to retain their shape when pushed out of a silicone tubing (Fig. 1). A gel of 26 mg/ml had an elasticity of 1000 to 2000 Pa, measured as Young's modulus by atomic force microscopy, which is comparable to a 0.4% agarose gel.

Remarkably, the gels remained stable and unaltered in appearance when the temperature was raised to 95°C. Apparently, the increased strength of hydrophobic interactions at increased temperature (*18*) can compensate for the increase of gel-destabilizing thermal motions. The gels, however, were readily dissolved by chaotropic agents such as 6 M guanidinium chloride; this result suggested that gel formation originated indeed from noncovalent, reversible interchain connections.

We next mutated each of the 55 phenylalanines (F) within the FG context to serines (S) (fig. S2). If hydrophobic cross-linking between fsFG repeats occurred, then it should be abolished in this mutant. Indeed, this F→S mutated repeat domain showed no signs of gelling and remained liquid in aqueous solution (Fig. 1), even when the concentration of repeats was raised to 100 mg gel/ml. Thus,

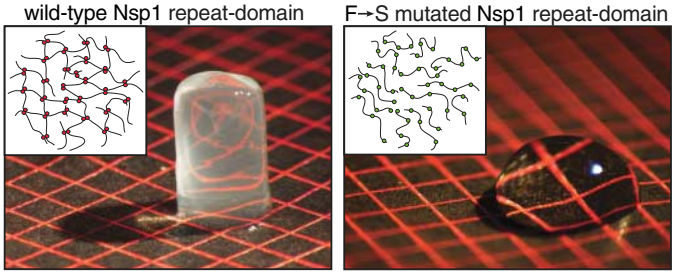


Fig. 1. FG repeats can form an elastic hydrogel in aqueous solution. (**Left**) An aqueous solution with 26 mg/ml wild-type fsFG-repeat domain from Nsp1p (400 μM) was filled into a silicon tubing, where it completed gelling. The formed gel was pushed out of the tubing by gentle pressure, placed onto a patterned support (1 square = 1.4 mm), and photographed. Note that the pattern shows clearly through this transparent gel. Inset illustrates how interactions between the hydrophobic clusters (shown in red) cross-link the repeat domains into a hydrogel. (**Right**) The F→S mutated repeat domain (see fig. S2) remained liquid after identical treatment.

[1]Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), INF 282, D-69120 Heidelberg, Germany. [2]Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg 11, D-37077 Göttingen, Germany. [3]Institut für Physikalische Chemie, Universität Heidelberg, INF253, D-69120 Heidelberg, Germany. [4]Max-Planck-Institut für Metallforschung, Heisenbergstrasse 3, D-70569 Stuttgart, Germany.

*To whom correspondence should be addressed. E-mail: dg@zmbh.uni-heidelberg.de

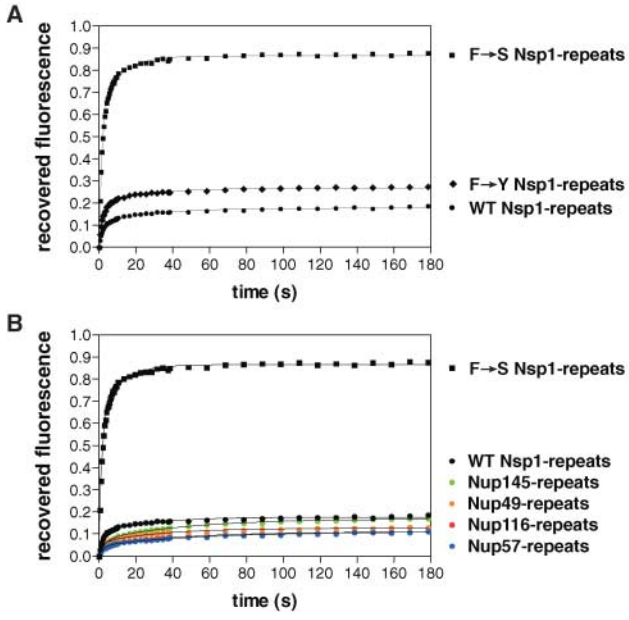


Fig. 2. (**A**) The wild-type Nsp1-repeat hydrogel interacts strongly with the wild-type and F→Y mutated repeat domain, but not with the F→S mutated version. A hydrogel with 400 μM unlabeled wild-type Nsp1 fsFG repeats and 1 μM of indicated fluorescently labeled repeat domains was formed as described in Fig. 1. The mobility of the labeled repeats was determined by FRAP. This method measures the diffusion-driven exchange of bleached molecules with unbleached ones. Photobleached regions of wild-type or F→Y mutated repeats recovered only a small fraction of their initial fluorescence; these repeats were essentially immobile and firmly incorporated into the gel. The F→S mutated version, in contrast, diffused freely inside the gel and recovered fluorescence almost completely. (**B**) GLFG repeats coassemble with Nsp1 fsFG repeats into a hydrogel. Hydrogels with 400 μM unlabeled wild-type Nsp1 fsFG-repeat domain and 1 μM of the indicated fluorescently labeled nucleoporin repeat domain were prepared and analyzed by FRAP. The GLFG repeats from Nup49p, Nup57p, Nup116p, or Nup145p were essentially immobile, which indicated strong heterotypic interactions with the fsFG repeats from Nsp1p.

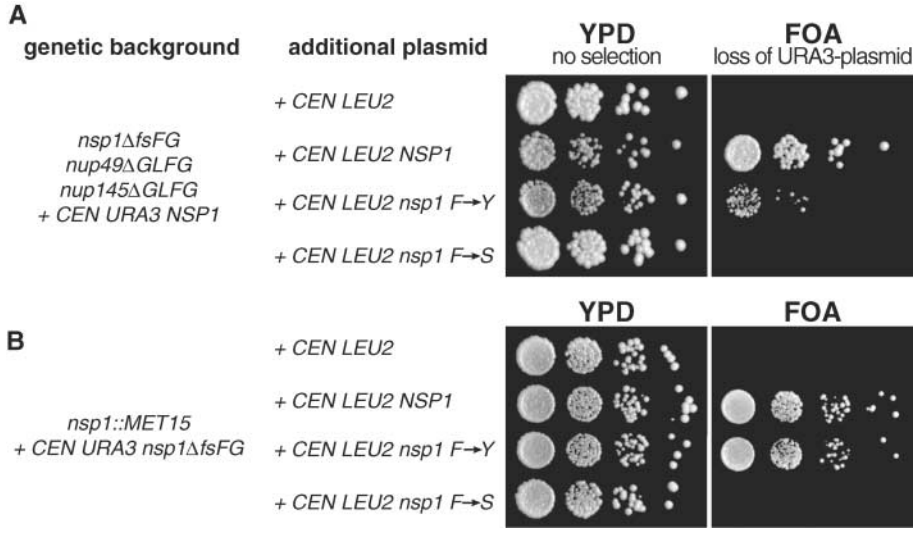


Fig. 3. In vivo effects of F→S and F→Y mutated FG-repeat domains. (**A**) A yeast strain lacking the GLFG-repeat regions of Nup49p and Nup145p, as well as the fsFG-repeat region of Nsp1p, was kept viable by expressing full-length Nsp1p from a low-copy *ARS/CEN* (*CEN*) plasmid harboring a *URA3* marker. The strain was subsequently transformed with indicated *CEN* plasmids carrying *LEU2* as a selectable marker and cultivated for several generations in synthetic complete (SC) medium lacking leucine. To select for cells that lost the original *CEN URA3* expression plasmid for wild-type Nsp1p, equal numbers of cells were plated in 10-fold serial dilutions on plates containing 5-fluoroorotic acid (FOA), which is toxic in the presence of a functional *URA3* gene. The total number of viable cells was determined in parallel by plating cells by the same serial dilution on full medium (YPD). The empty *CEN LEU2* plasmid used as the negative control did not support growth on FOA, which showed that the fsFG repeats of Nsp1p are essential in this genetic background (*19*, *20*). Cells were viable on FOA if the remaining *CEN LEU2* plasmid encoded wild-type Nsp1p. Cells expressing Nsp1p with F→Y mutated repeats were also viable on FOA, but showed a growth defect. Expression of Nsp1p with F→S mutated repeats did not allow for growth on FOA. (**B**) A yeast strain with a genomic deletion of *NSP1* was complemented by a *CEN URA3* plasmid that allowed expression of Nsp1p without fsFG repeats (nsp1ΔfsFG). The FOA test showed that this *URA3* plasmid could be lost, provided wild-type Nsp1p or Nsp1p with F→Y mutated repeats was expressed from the *CEN LEU2* plasmid. The strain expressing Nsp1p with F→S mutated repeats was not viable on FOA, i.e., it died upon loss of the nsp1ΔfsFG-encoding *URA3* plasmid.

inter-repeat contacts between phenylalanines caused the gelling of the wild-type Nsp1 fsFG-repeat domain.

By fluorescence recovery after photobleaching (FRAP), we next established that a fluorescently labeled wild-type Nsp1 fsFG-repeat domain was nearly immobile within the wild-type fsFG-repeat hydrogel (Fig. 2A). In contrast, a fluorescently labeled F→S mutated repeat domain showed no interaction and diffused freely within the wild-type gel (Fig. 2A), which confirmed that the phenyl groups act as mediators of the interactions.

As a next step, we probed heterotypic interactions between the FXFG and GLFG repeats and found that the GLFG-repeat domains from Nup49p, Nup57p, Nup145p, and Nup116p became firmly incorporated into the Nsp1 fsFG-repeat hydrogel (Fig. 2B). Considering that FG, FXFG, and GLFG repeats account for ~12% of the total NPC mass (*15*), that all repeat types tested so far could engage into inter-repeat interactions (Fig. 1 and Fig. 2B), and that their estimated local concentration at NPCs of ~100 mg/ml (*7*) exceeds the critical concentration for gelling (~8 mg gel per milliliter), we propose that extensive regions of the NPC and possibly the entire permeability barrier are organized in such a hydrogel structure.

Having established that the wild-type Nsp1 repeats formed a hydrogel in vitro, while the F→S mutated repeats did not, we wanted to elucidate the effect of this mutation on NPC function in vivo. Because of redundancy, none of the individual FXFG-repeat or GLFG-repeat domains is essential for NPC function in the yeast *Saccharomyces cerevisiae*, and only combined deletions are lethal (*19*). We thus chose genetic backgrounds where the Nsp1 fsFG-repeat domain is essential for viability (*19*) and found that full-length Nsp1p with an F→S mutated repeat domain ("Nsp1p F→S") was unable to substitute for wild-type Nsp1p (Fig. 3A and fig. S3).

We then repeated the experiment in a wild-type genetic background, where the fsFG repeats of Nsp1p are nonessential (*19*, *20*). Here, expression of the Nsp1 rod domain (which lacks the fsFG-repeat domain) was sufficient to rescue the lethal chromosomal *NSP1* deletion (*20*) (Fig. 3B). As expected, the lethal phenotype of the complete *NSP1* deletion was also rescued by expressing wild-type full-length Nsp1p. In contrast, Nsp1p with F→S mutated repeats failed to complement the *NSP1* deletion (Fig. 3B), even though the mutant protein was expressed, stable (fig. S4), and able to bind its interaction partners Nup49p and Nup57p (*21*) (fig. S5), as well as to assemble into NPCs (fig. S6). Thus, although deletion of the fsFG repeats from Nsp1p was tolerated, replacement of all copies of Nsp1p in NPCs by the F→S mutated version was lethal. In other words, the permeability barrier can tolerate permuting the anchoring points for FG repeats at the NPC

structure (*22*), as well as a significant reduction in mass (*19*), but not a major change toward a more hydrophilic composition.

The F→S mutated Nsp1 repeats failed to form a hydrogel (Figs. 1 and 2) as well as to bind NTRs (Fig. 4A). To distinguish which of these failures might cause the described lethal effect in vivo, we designed a less drastic mutant, where the phenylalanines (F) within the Nsp1 fsFG-repeat domain had been mutated to tyrosines (Y). The F→Y mutated repeats also failed to bind NTRs (Fig. 4A), probably because the additional OH group at the phenyl ring cannot be accommodated into the FXFG binding pockets of the NTRs (*8*, *9*, *23*, *24*). Remarkably, *nsp1* with F→Y mutated repeats fully complemented the *NSP1* deletion in a wild-type background (Fig. 3B) and even showed complementation in several of the genetic backgrounds, where the fsFG repeats of Nsp1p are essential (Fig. 3A and fig. S3). Thus, the F→Y mutation only abolished NTR binding, but apparently left other interactions intact that are essential, but redundant with other repeat domains.

Therefore, we tested the behavior of the F→Y mutant repeats inside the wild-type fsFG-repeat gel and observed by FRAP a low mobility of the F→Y mutant, i.e., a clear binding between F→Y mutant and wild-type Nsp1 repeats (Fig. 2A). Perhaps the underlying interactions are not purely hydrophobic, but involve also "aromatic" interactions: π stacking and certain geometries of T-shape interactions of the π-electron systems could occur, not only between two phenyl rings, but also between a phenyl and the hydroxyphenyl group of tyrosine, or even between two hydroxyphenyl groups. Indeed, the F→Y mutated repeat domain also formed a homotypic hydrogel (Fig. 4B).

FG-rich nucleoporin repeats are essential for viability (*18*) and engage in two known kinds of interactions: binding of NTRs (*6*, *25*) and hydrogel formation that arises through inter-repeat contacts (this study). If only the NTR binding, but not the inter-repeat contacts, represented essential functions, then the F→S and the F→Y mutant should both behave as if they were complete loss-of-function mutants. We observed, however, a different scenario, namely that the F→Y mutant could rescue, in certain genetic backgrounds, an otherwise lethal deletion of FG repeats. The simplest explanation for these data are that inter-repeat contacts and, hence, hydrogel formation are also required for NPC function and viability.

How could NTRs cross a permeability barrier containing FG repeats that are F→Y mutated and, hence, unable to bind NTRs? NTRs should still be able to locally dissolve the reversible Phe-Tyr cross-links by binding the "wild-type side." However, they should fail to break Tyr-Tyr contacts, which, in turn, could explain why in some genetic backgrounds the mutation is not tolerated (fig. S3).

An analogous consideration could explain also a puzzling set of previous findings: Even though NTRs are probably all optimized to cross the permeability barrier rapidly, they differ in their affinity for the different types of FG repeats (*26*), and they are differently affected by genetic deletions of individual FG-repeat domains (*19*). NTRs should indeed have problems penetrating homotypic hydrogels of those FG repeats to which these NTRs bind too weakly. Mixed gels formed by heterotypic interactions, however, should pose less of a problem and should constitute a more robust system, simply because an NTR could locally dissolve the hydrophobic contact between two different repeat units by binding to any of them.

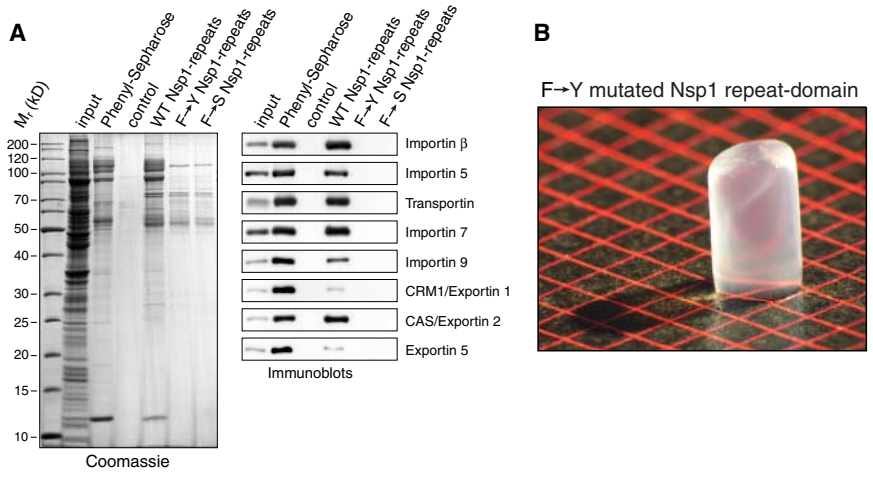


Fig. 4. (**A**) Mutation of the phenylalanines within the Nsp1-repeat domain to serines or tyrosines prevents binding of NTRs. A hypotonic extract from human HeLa cells was bound to indicated forms of immobilized Nsp1-repeat domains. Bound fractions were analyzed by SDS–polyacrylamide electrophoresis (SDS-PAGE), followed by either Coomassie staining or immunoblotting with antibodies specific for indicated importins and exportins. Phenyl-Sepharose served as positive control for NTR binding (*12*). Importins and exportins typically migrate between 90 and 135 kD, NTF2 at 12 kD. Nsp1 repeats were covalently bound via an engineered C-terminal cysteine to iodoacetamide-Sepharose 4B. No protein had been coupled to the negative control beads. Nonreacted iodoacetamide groups had been quenched with β-mercaptoethanol. Further tests revealed that yeast NTRs such as importin β (Kap95p) or Ntf2p behaved as the human ones. (**B**) F→Y mutated Nsp1-repeat domains also form a homotypic hydrogel. F→Y mutated repeat domains (400 μM) were treated as the wild-type repeat domain in Fig. 1; however, in order to avoid oxidation of the tyrosines, the handling was under argon, and β-mercaptoethanol (50 mM) was added as a scavenger. The resulting hydrogel had elastic properties similar to those of the wild-type fsFG-repeat gel, although it was not as transparent and, thus, was less homogeneous.

References and Notes

1. D. Görlich, U. Kutay, *Annu. Rev. Cell Dev. Biol.* **15**, 607 (1999).
2. I. G. Macara, *Microbiol. Mol. Biol. Rev.* **65**, 570 (2001).
3. L. F. Pemberton, B. M. Paschal, *Traffic* **6**, 187 (2005).
4. K. Weis, *Curr. Opin. Cell Biol.* **14**, 328 (2002).
5. A. Radu, M. S. Moore, G. Blobel, *Cell* **81**, 215 (1995).
6. M. K. Iovine, J. L. Watkins, S. R. Wente, *J. Cell Biol.* **131**, 1699 (1995).
7. R. Bayliss *et al.*, *J. Mol. Biol.* **293**, 579 (1999).
8. R. Bayliss, T. Littlewood, M. Stewart, *Cell* **102**, 99 (2000).
9. J. Bednenko, G. Cingolani, L. Gerace, *J. Cell Biol.* **162**, 391 (2003).
10. M. P. Rout, S. R. Wente, *Trends Cell Biol.* **4**, 357 (1994).
11. K. Ribbeck, D.Görlich, *EMBO J.* **20**, 1320 (2001).
12. K. Ribbeck, D.Görlich, *EMBO J.* **21**, 2664 (2002).
13. D. P. Denning, S. S. Patel, V. Uversky, A. L. Fink, M. Rexach, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2450 (2003).
14. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; and X, any amino acid. In FXFG repeats, X is preferentially S, T, G, A, or N.
15. M. P. Rout *et al.*, *J. Cell Biol.* **148**, 635 (2000).
16. E. C. Hurt, *EMBO J.* **7**, 4323 (1988).
17. Materials and methods are available as supporting material on *Science* Online.
18. D. Chandler, *Nature* **437**, 640 (2005).
19. L. A. Strawn, T. Shen, N. Shulga, D. S. Goldfarb, S. R. Wente, *Nat. Cell Biol.* **6**, 197 (2004).
20. U. Nehrbass *et al.*, *Cell* **61**, 979 (1990).
21. P. Grandi, N. Schlaich, H. Tekotte, E. C. Hurt, *EMBO J.* **14**, 76 (1995).
22. B. Zeitler, K. Weis, *J. Cell Biol.* **167**, 583 (2004).
23. S. Fribourg, I. C. Braun, E. Izaurralde, E. Conti, *Mol. Cell* **8**, 645 (2001).
24. R. Bayliss *et al.*, *EMBO J.* **21**, 2843 (2002).
25. A. Radu, G. Blobel, M. S. Moore, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1769 (1995).
26. I. Ben-Efraim, L. Gerace, *J. Cell Biol.* **152**, 411 (2001).

27. We thank S. Wente for discussions and yeast strains; E. Hurt for the donation of plasmids; J. Spatz for support; T. A. Rapoport, T. Güttler, B. Hülsmann, V. C. Cordes, and J. E. Curtis for critical reading of the manuscript, as well as the Deutsche Forschungsgemeinschaft (SFB 638) and the Max-Planck-Gesellschaft for financial support.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/815/DC1
Materials and Methods
Figs. S1 to S6
References

14 July 2006; accepted 15 September 2006
10.1126/science.1132516

Protrudin Induces Neurite Formation by Directional Membrane Trafficking

Michiko Shirane and Keiichi I. Nakayama*

Guanosine triphosphatases of the Rab family are key regulators of membrane trafficking, with Rab11 playing a specific role in membrane recycling. We identified a mammalian protein, protrudin, that promoted neurite formation through interaction with the guanosine diphosphate (GDP)–bound form of Rab11. Phosphorylation of protrudin by extracellular signal–regulated kinase (ERK) in response to nerve growth factor promoted protrudin association with Rab11-GDP. Down-regulation of protrudin by RNA interference induced membrane extension in all directions and inhibited neurite formation. Thus, protrudin regulates Rab11-dependent membrane recycling to promote the directional membrane trafficking required for neurite formation.

The molecular mechanisms that underlie neurite formation include both cytoskeletal remodeling and membrane trafficking (*1*, *2*). During this process, membrane components are transported in a directional manner within the cell by a membrane recycling system, resulting in expansion of the cell surface area in the region of neurite formation (*3*, *4*). Rab11 regulates membrane recycling back to the plasma membrane and constitutive exocytosis (*5*–*14*).

We isolated protrudin as a protein that interacts with FKBP38, a multifunctional membranous chaperone (*15*). Protrudin contains a Rab11 binding domain (RBD11), a guanosine diphosphate (GDP) dissociation inhibitor (GDI) consensus sequence (*16*), two hydrophobic domains (HP-1 and HP-2), an FFAT motif (an endoplasmic reticulum–targeting signal) (*17*), a coiled-coil domain, and a FYVE domain (a phosphoinositide binding domain) (fig. S1A) (*18*). These characteristics suggested that protrudin might function in membrane trafficking, especially in membrane recycling.

Expression of FLAG-protrudin in HeLa cells resulted in the generation of long processes (Fig. 1A). Processes were observed in 5 to 30% of transfected cells. We also monitored process formation by time-lapse video microscopy in living HeLa cells expressing enhanced green fluorescent protein (EGFP)–protrudin (Fig. 1B and movie S1). The processes appeared to grow in a manner similar to that of neurites with ruffling lamellipodia. In rat hippocampal neurons, expression of the exogenous protrudin promoted neurite extension in the neurons (Fig. 1C and fig. S2). We next analyzed the effects of depletion of endogenous protrudin by RNA interference (RNAi) in the rat pheochromocytoma PC12 cell line, which extends neurites in response to nerve growth factor (NGF) (Fig. 1D). Depletion of protrudin resulted in inhibition of NGF-induced neurite outgrowth and triggered spreading of the plasma membrane in all directions (Fig. 1, E and F). Thus, protrudin is necessary for neurite formation and may function by regulating the direction of membrane sorting.

In mouse, protrudin was most abundant in the cerebrum and the cerebellum (Fig. 2A). In mouse primary neurons in the early stage of differentiation, protrudin was localized to the pericentrosomal region and neurites (Fig. 2B). At later stages, it was detected in both dendrites and axons and was concentrated at the growth cones (Fig. 2C). We next examined the effect of NGF on the subcellular distribution of protrudin in PC12 cells (Fig. 2D). In the absence of NGF, protrudin was distributed diffusely throughout the cell body. In cells exposed to NGF for 6 hours, protrudin was markedly concentrated in the pericentrosomal region, where it colocalized with the recycling endosome marker Rab11 (fig. S3, A to C). After culture with NGF for 24 hours, protrudin was distributed from the pericentrosomal region to the tips of neurites (Fig. 2D). The overall abundance of protrudin was not affected by NGF (fig. S3D). Immuno–electron microscopic analysis revealed that protrudin appears to be associated with vesicle membranes (fig. S3E). Thus, protrudin undergoes a dynamic redistribution during neuritogenesis.

Although protrudin contains a consensus sequence for an RBD11, the amino acids required for interaction with the GTP-bound form of Rab11 are not conserved in this domain of protrudin (*19*) (Fig. 3A). The amino acid sequence of the RBD11 of protrudin is similar to the corresponding sequences of GDI-α and GDI-β, both of which interact with the GDP-

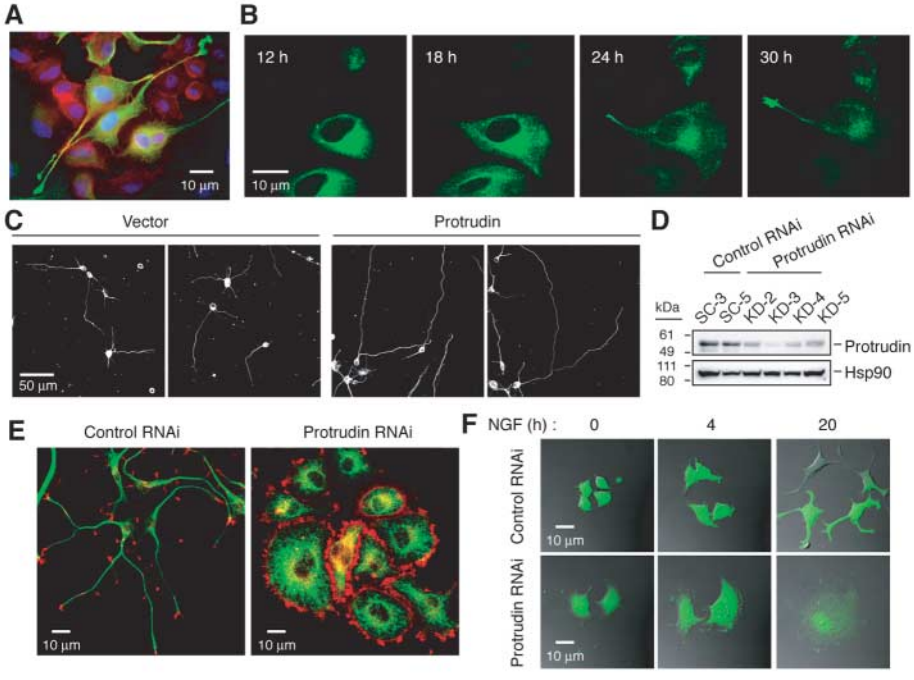


Fig. 1. Protrudin induces outgrowth of cellular projections. (**A**) HeLa cells were transfected with a FLAG-protrudin vector and stained with anti-FLAG (green), phalloidin (red), and Hoechst 33258 (WAKO, Osaka, Japan) (blue). (**B**) HeLa cells were transfected with EGFP-protrudin, and the fluorescence was monitored by video (movie S1). (**C**) Rat hippocampal neurons were transfected with FLAG-protrudin or the empty vector and stained with anti–βIII-tubulin. (**D**) PC12 cells were transfected with short hairpin RNA (shRNA) vectors specific for protrudin mRNA (KD-2, -3, -4, and -5) or control (SC-3 and -5) and subjected to immunoblotting with anti-protrudin and anti–heat shock protein 90 (anti-Hsp90). (**E**) PC12 cells were transfected with control (SC-3) or protrudin (KD-3) shRNA vectors, stimulated with NGF, and stained with anti–α-tubulin (green) and phalloidin (red). (**F**) PC12 cells were transfected with a vector encoding Venus and either control (SC-3) or protrudin (KD-3) shRNA, stimulated with NGF, and examined for Venus fluorescence.

Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan. Precursory Research for Embryonic Science and Technology and Core Research for Evolutional Science and Technology, Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan.

*To whom correspondence should be addressed. E-mail: nakayak1@bioreg.kyushu-u.ac.jp

bound form of many Rab proteins (*20*, *21*). We thus examined whether protrudin actually interacted with Rab11. Endogenous protrudin did indeed interact with endogenous Rab11 in PC12 cells, and this interaction was enhanced by NGF treatment (Fig. 3B). FLAG-protrudin also interacted with hemagglutinin (HA)–tagged Rab11 in human embryonic kidney (HEK293T) cells (Fig. 3C). Protrudin preferentially interacted with the GDP-bound form of Rab11 [as represented by the GTP binding–deficient mutant Ser^{25}→Asn^{25} (S25N)] rather than with the GTP-bound form [as represented by the GTPase-deficient mutant Gln^{70}→Leu^{70} (Q70L)] (*22*, *23*) (Fig. 3D). The RBD11 of the Rab11 effector FIP2 interacted specifically with Rab11-GTP, whereas the RBD11 of protrudin specifically interacted with Rab11-GDP (fig. S4A). Reciprocal co-immunoprecipitation analysis as well as an in vitro pull-down assay confirmed that protrudin preferentially interacts with Rab11-GDP (fig. S4, B and C). We also generated a series of deletion mutants of protrudin (fig. S1B) and tested them for the ability to bind Rab11. Rab11(S25N) interacted with the N1 mutant but not with the ΔRBD-N1 mutant (fig. S4D). Furthermore, a protrudin deletion mutant lacking the RBD11 (ΔRBD in fig. S1B) could not restore neurite formation in PC12 cells depleted of protrudin by RNAi (fig. S5A) or induce process formation in HeLa cells (fig. S5B). Thus, protrudin preferentially binds to Rab11-GDP, and this association is required for neurite formation.

Sustained activation of the mitogen-activated protein kinase (MAPK) ERK accompanies NGF-dependent neurite extension in PC12 cells (*24*) (fig. S6A). PC12 cells expressing a constitutively active form of MEK1 [MEK1(SDSE)], which activated ERK, manifested neurite formation in the absence of NGF (fig. S6, B and C). Protrudin and MEK1(SDSE) showed a synergistic effect on process formation in HeLa cells, whereas a kinase-negative form of MEK1 [MEK1(K97S)] antagonized the process-forming activity of protrudin (fig. S6D). Protrudin contains six potential ERK phosphorylation sites as well as two consensus ERK binding (ERK D) domains (fig. S1C). Phosphorylation of protrudin was increased by NGF treatment or MEK1(SDSE) expression, but not by MEK1(K97S) expression, in PC12 cells (Fig. 3E). Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) also suggested that protrudin is phosphorylated by ERK in response to NGF (Fig. 3F). MEK1(K97S) expression inhibited the NGF-induced shift, suggesting that ERK activation is essential for the phosphorylation of protrudin elicited by NGF.

We examined the effect of ERK activation on protrudin-Rab11 interaction in HEK293T cells. Expression of MEK1(SDSE) enhanced the interaction between protrudin and Rab11(S25N) (Fig. 3G). At high amounts of MEK1(SDSE) expression, both wild-type and S25N forms of Rab11 interacted substantially with protrudin (fig. S7, A and B). Furthermore, protrudin mutants that lack some of potential ERK phosphorylation sites (P-mut-1 and -4) or both intact ERK D domains (D-mut) (fig. S1C) showed markedly reduced affinities for Rab11 compared with the affinities of wild-type protrudin (Fig. 3H and fig. S7C). Replacement of potential ERK phosphorylation sites in other combinations (P-mut-2 and -3) did not affect the binding (fig. S7D). Thus, phosphorylation of protrudin at multiple sites in response to NGF-ERK signaling promotes its interaction with Rab11.

We next investigated the effect of Rab11 on the morphology of PC12 cells. The morphology of cells expressing Rab11(Q70L) appeared nearly identical to that of cells depleted of endogenous protrudin (Fig. 4A). Conversely, the phenotype conferred by expression of Rab11(S25N) was similar to that conferred by expression of protrudin. These experiments were combined to examine the genetic relation between protrudin and Rab11. Expression of Rab11(S25N) induced neurite formation in cells depleted of endogenous protrudin (Fig. 4B), whereas overexpression of protrudin had no effect on the morphology of cells expressing Rab11(Q70L) (Fig. 4C). In addition, expression of Rab11(Q70L) inhibited process formation induced by protrudin in HeLa cells (movie S2). Thus, protrudin is indeed an upstream inhibitor of Rab11 function.

To investigate the role of the protrudin-Rab11 system in directed membrane traffic, we observed the transport of NgCAM, a cell adhesion molecule that is delivered to the somatodendritic plasma membrane and then transported to the axonal plasma membrane via recycling endosomes (*25*–*27*). Chicken NgCAM expressed in PC12 cells initially accumulated, presumably in the endoplasmic reticulum–Golgi compartment, and was subsequently transported to the plasma membrane of some neurites, but not to that of the soma, in the presence of NGF (Fig. 4D and fig. S8). In contrast, cells depleted of protrudin by RNAi or those expressing Rab11(Q70L) exhibited prominent surface

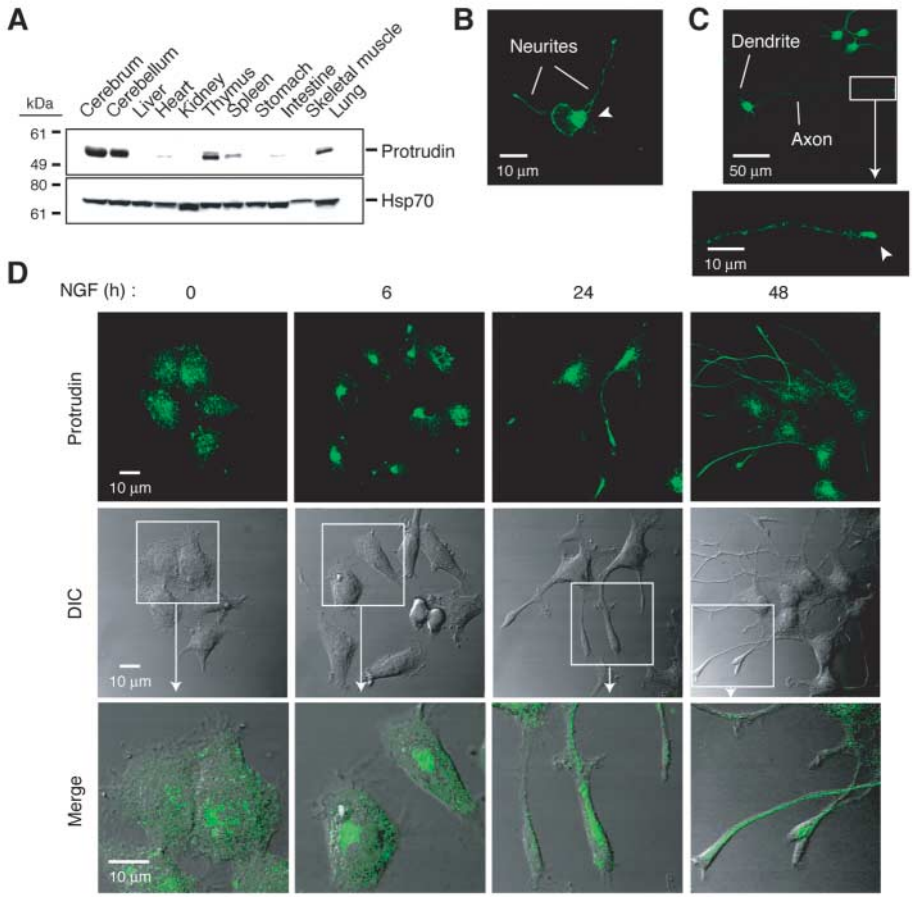


Fig. 2. Subcellular redistribution of protrudin during neuritogenesis. (**A**) Tissue distribution of protrudin in mice. (**B** and **C**) Cells isolated from embryonic mouse cerebral cortex were cultured for 1 (B) or 3 (C) days and stained with anti-protrudin (green). Arrowhead in (B) indicates the pericentrosomal region. The boxed region in (C) is shown at higher magnification in the bottom image; the arrowhead indicates the growth cone. (**D**) PC12 cells were incubated with NGF and stained with anti-protrudin (top, green). Differential interference contrast (DIC) images are shown in the middle, and the boxed regions are shown at bottom merged with protrudin.

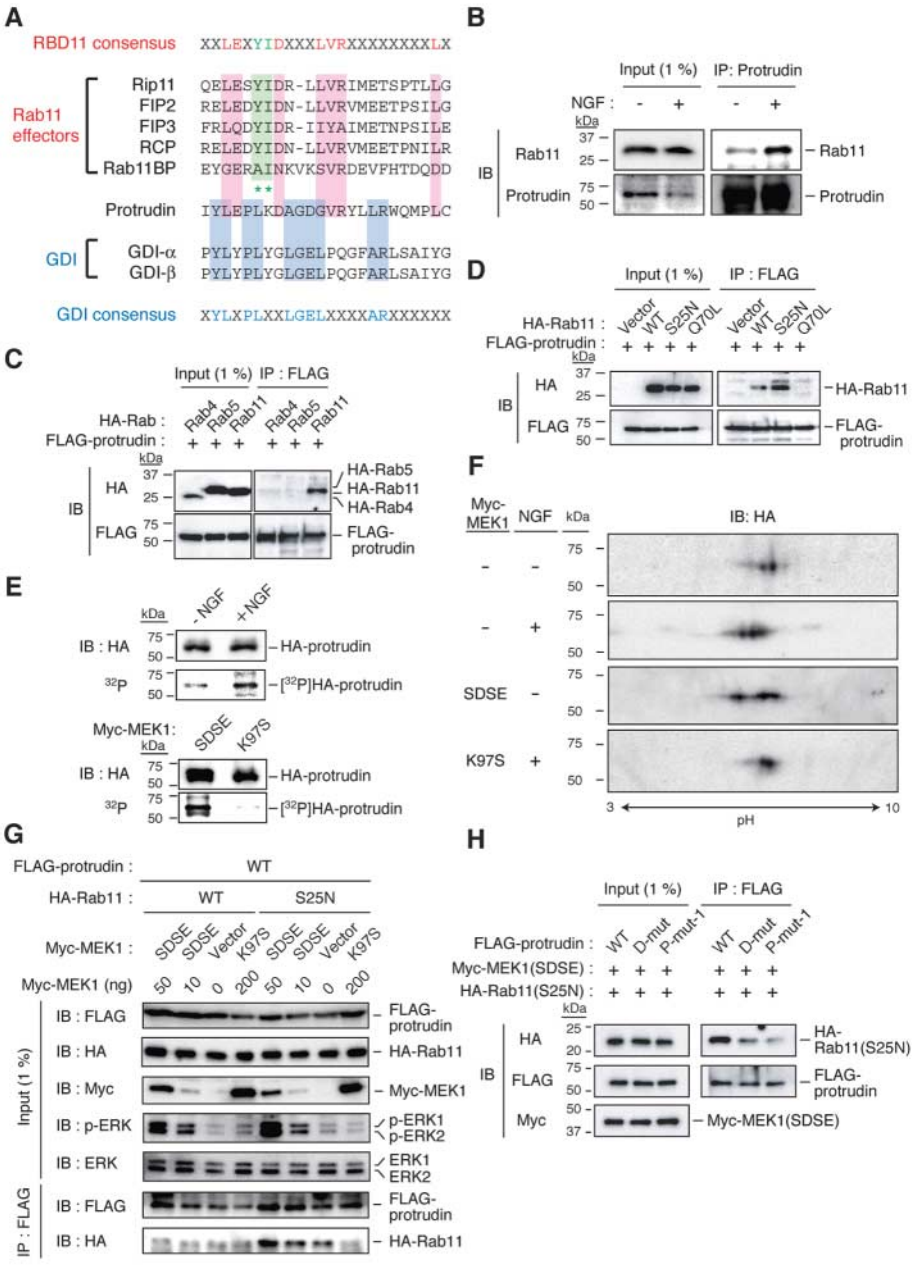


Fig. 3. Interaction of protrudin with Rab11. (**A**) Alignment of the amino acid sequence (*17*) of the RBD11 of protrudin with the Rab11 effector proteins and GDIs. Conserved residues are shaded in red or blue. Asterisks denote critical residues for interaction with Rab11-GTP (shaded in green). (**B**) Immunoprecipitation (IP) of PC12 lysates with anti-protrudin and immunoblotting (IB) with anti-Rab11 and -protrudin. (**C**) FLAG-protrudin expressed with HA-Rab4, -Rab5, or -Rab11 in HEK293T cells was immunoprecipitated and subjected to IB with anti-HA and -FLAG. (**D**) FLAG-protrudin expressed with either wild-type (WT) or mutant (S25N or Q70L) HA-Rab11 in HEK293T cells was immunoprecipitated and subjected to IB with anti-HA and -FLAG. (**E**) PC12 cells expressing HA-protrudin were metabolically labeled with [^{32}P]orthophosphate in the absence or presence of NGF (top). Alternatively, PC12 cells expressing HA-protrudin and Myc-MEK1 (SDSE or K97S) were similarly labeled (bottom). Cell lysates were subjected to IP with anti-HA and to autoradiography or to IB with anti-HA. (**F**) PC12 cells expressing HA-protrudin as well as Myc-MEK1 (SDSE or K97S) were incubated with or without NGF. Cell lysates were subjected to IP with anti-HA and to 2D-PAGE followed by IB with anti-HA. (**G**) HEK293T cells transfected with FLAG-protrudin, HA-Rab11 (WT or S25N), and Myc-MEK1 (SDSE or K97S) were subjected to IP with anti-FLAG and to IB with anti-FLAG, -HA, -Myc, -ERK, and -phospho-ERK (p-ERK). (**H**) HEK293T cells transfected with FLAG-protrudin (WT, D-mut, or P-mut-1), Myc-MEK1(SDSE), and HA-Rab11(S25N) were subjected to IP with anti-FLAG and to IB with anti-HA, -FLAG, and -Myc.

expression of NgCAM on the soma (Fig. 4D). Thus, protrudin is essential for directional membrane trafficking to neurites.

Our data indicate that the protrudin-Rab11 system is an important determinant of the direction of membrane trafficking and neurite formation. It was recently shown that ZFYVE27 (a synonym of protrudin) is mutated in a German family with an autosomal dominant form of hereditary spastic paraplegia (AD-HSP), which is characterized by selective degeneration of axons (*28*). The phenotype of the affected individuals was similar to that of patients with AD-HSP caused by mutation of spastin, a protein implicated in neuronal vesicular cargo trafficking. Protrudin is thought to interact with spastin via its FYVE domain in the COOH-terminal region of the protein. This genetic evidence supports our conclusion that protrudin plays a central role in membrane trafficking in neurons. Rab11, protrudin, and spastin may together constitute a system for the regulation of vesicular transport in neurons, and impairment of this system may be responsible for the pathogenesis of AD-HSP.

References and Notes

1. L. S. Goldstein, Z. Yang, *Annu. Rev. Neurosci.* **23**, 39 (2000).
2. A. M. Craig, G. Banker, *Annu. Rev. Neurosci.* **17**, 267 (1994).
3. A. H. Futerman, G. A. Banker, *Trends Neurosci.* **19**, 144 (1996).
4. B. L. Tang, *J. Neurochem.* **79**, 923 (2001).
5. R. Behnia, S. Munro, *Nature* **438**, 597 (2005).
6. R. Jahn, T. Lang, T. C. Sudhof, *Cell* **112**, 519 (2003).
7. J. S. Bonifacino, B. S. Glick, *Cell* **116**, 153 (2004).
8. M. Zerial, H. McBride, *Nat. Rev. Mol. Cell Biol.* **2**, 107 (2001).
9. F. R. Maxfield, T. E. McGraw, *Nat. Rev. Mol. Cell Biol.* **5**, 121 (2004).
10. S. Urbe, L. A. Huber, M. Zerial, S. A. Tooze, R. G. Parton, *FEBS Lett.* **334**, 175 (1993).
11. W. Chen, Y. Feng, D. Chen, A. Wandinger-Ness, *Mol. Biol. Cell* **9**, 3241 (1998).
12. G. Emery *et al.*, *Cell* **122**, 763 (2005).
13. S. Wu, S. Q. Mehta, F. Pichaud, H. J. Bellen, F. A. Quiocho, *Nat. Struct. Mol. Biol.* **12**, 879 (2005).
14. R. Prekeris, J. Klumperman, R. H. Scheller, *Mol. Cell* **6**, 1437 (2000).
15. M. Shirane, K. I. Nakayama, *Nat. Cell Biol.* **5**, 28 (2003).
16. J. R. Junutula *et al.*, *J. Biol. Chem.* **279**, 33430 (2004).
17. S. E. Kaiser *et al.*, *Structure* **13**, 1035 (2005).
18. D. J. Gillooly, A. Simonsen, H. Stenmark, *Biochem. J.* **355**, 249 (2001).
19. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
20. I. Schalk *et al.*, *Nature* **381**, 42 (1996).
21. Y. An *et al.*, *Structure* **11**, 347 (2003).
22. O. Ullrich, S. Reinsch, S. Urbe, M. Zerial, R. G. Parton, *J. Cell Biol.* **135**, 913 (1996).
23. S. Pasqualato *et al.*, *J. Biol. Chem.* **279**, 11480 (2004).
24. C. J. Marshall, *Cell* **80**, 179 (1995).
25. B. Sampo, S. Kaech, S. Kunz, G. Banker, *Neuron* **37**, 611 (2003).
26. D. Wisco *et al.*, *J. Cell Biol.* **162**, 1317 (2003).
27. E. Anderson *et al.*, *J. Cell Biol.* **170**, 595 (2005).
28. A. U. Mannan *et al.*, *Am. J. Hum. Genet.* **79**, 351 (2006).

29. We thank T. Takenawa, M. Takeichi, H. Togashi, Y. Takai, A. Kikuchi, A. Nakano, K. Hanada, H. Sumimoto, Y. Fukui, K. Mihara, Y. Fujiki, M. Hirata, T. Natsume, T. Sasaki, T. Nishi, M. Kohjima, R. Takeya, T. Uemura, and Y. Gotoh

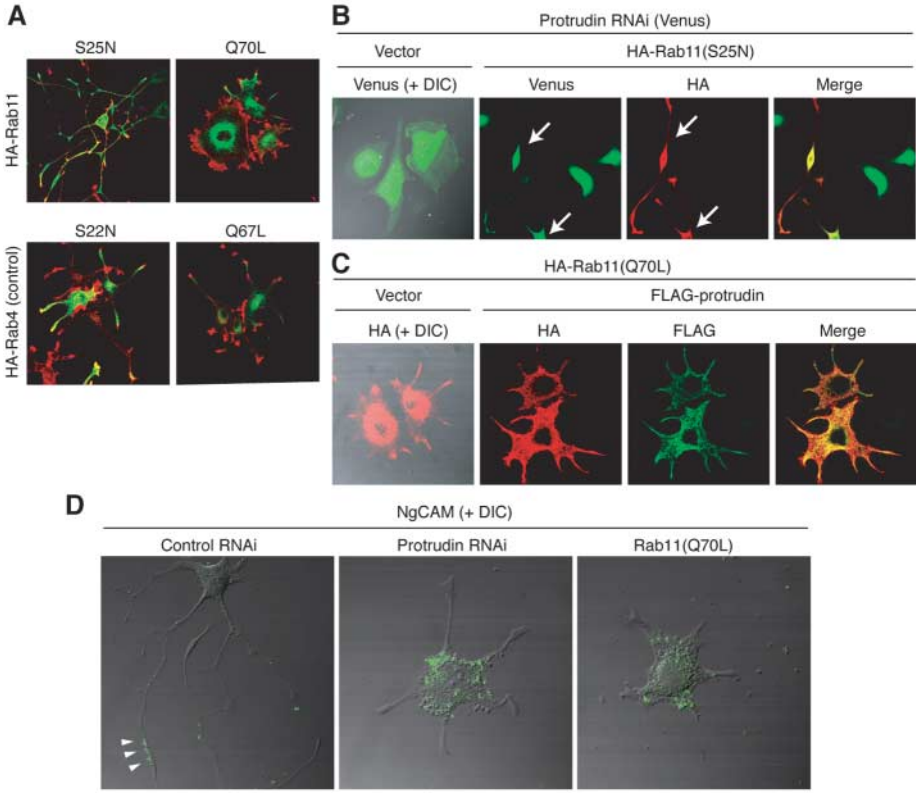


Fig. 4. Protrudin induces directional membrane extension through regulation of vesicular traffic. (**A**) PC12 cells expressing HA-Rab11 (S25N or Q70L) or HA-Rab4 (S22N or Q67L) were stimulated with NGF and stained with anti-HA (green) and phalloidin (red). (**B**) PC12 cells transfected with a vector encoding both protrudin shRNA and Venus (green) were also transfected with a HA-Rab11(S25N) vector, incubated with NGF, and stained with anti-HA (red). Arrows indicate cells expressing both Venus (protrudin shRNA) and HA-Rab11(S25N). (**C**) PC12 cells were transfected with HA-Rab11(Q70L) and FLAG-protrudin, stimulated with NGF, and stained with anti-HA (red) and anti-FLAG (green). (**D**) PC12 cells transiently infected with a retrovirus encoding chicken NgCAM were transfected with control or protrudin shRNAs or with Rab11(Q70L). The surface expression of NgCAM was examined by immunostaining with anti-NgCAM (green) without cell permeabilization. Arrowheads indicate the prominent distribution of NgCAM on a neurite.

for discussion; M. Sasaki for electron microscopy; E. Nishida for MEK1 cDNAs; A. Miyawaki for Venus cDNA; P. Sonderegger and B. Winckler for NgCAM cDNA; T. Chano for the pCL-Ampho vector; N. Tanaka for PC12 cells; N. Nishimura, R. Mitsuyasu, F. Matsuzaki, K. Azuma, and K. Oyamada for technical assistance; and M. Kimura and A. Ohta for help in preparation of the manuscript. This work was supported by funds from JST to M.S. and K.I.N.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/818/DC1
Materials and Methods
Figs. S1 to S8
Movies S1 and S2

17 August 2006; accepted 21 September 2006
10.1126/science.1134027

Where Water Is Oxidized to Dioxygen: Structure of the Photosynthetic Mn_4Ca Cluster

Junko Yano,[1,2]* Jan Kern,[3]* Kenneth Sauer,[1,2] Matthew J. Latimer,[4] Yulia Pushkar,[1,2] Jacek Biesiadka,[5] Bernhard Loll,[5]† Wolfram Saenger,[5] Johannes Messinger,[6]‡ Athina Zouni,[3]‡ Vittal K. Yachandra[1]‡

The oxidation of water to dioxygen is catalyzed within photosystem II (PSII) by a Mn_4Ca cluster, the structure of which remains elusive. Polarized extended x-ray absorption fine structure (EXAFS) measurements on PSII single crystals constrain the Mn_4Ca cluster geometry to a set of three similar high-resolution structures. Combining polarized EXAFS and x-ray diffraction data, the cluster was placed within PSII, taking into account the overall trend of the electron density of the metal site and the putative ligands. The structure of the cluster from the present study is unlike either the 3.0 or 3.5 angstrom–resolution x-ray structures or other previously proposed models.

Oxygen, which makes up about 20% of Earth's atmosphere, comes mostly from photosynthesis that occurs in cyanobacteria, green algae, and higher plants (*1*). These organisms have within photosystem II (PSII) an oxygen-evolving complex (OEC), in which the energy of sunlight is used to oxidize water to molecular oxygen. The heart of the OEC is a cluster of four Mn atoms and one Ca atom (Mn_4Ca) connected by mono-μ-oxo, di-μ-oxo, and/or hydroxo bridges. The specific protein environment and one chloride ion are also essential for the water-splitting activity (*1*). During the oxidation of water, the OEC cycles through five different oxidation states, which are known as S_i states (where *i* ranges from 0 to 4), that couple the one-electron photochemistry of the PSII reaction center with the four-electron chemistry of water oxidation (*2*).

The structure of the Mn_4Ca cluster and its role in the mechanism of water oxidation have been investigated with the use of spectroscopic methods (*1*), especially electron paramagnetic resonance and electron nuclear double-resonance spectroscopy (*3*–*9*), x-ray spectroscopy (*10*), and Fourier transform infrared (FTIR) spec-

[1]Melvin Calvin Laboratory, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [2]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [3]Max-Volmer-Laboratorium für Biophysikalische Chemie, Technische Universität, Strasse des 17 Juni 135, D-10623 Berlin, Germany. [4]Stanford Synchrotron Radiation Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94305, USA. [5]Institut für Chemie und Biochemie/Kristallographie, Freie Universität, Takustrasse 6, D-14195 Berlin, Germany. [6]Max-Planck-Institut für Bioanorganische Chemie, Stiftstrasse 34–36, D-45470 Mülheim an der Ruhr, Germany.

*These authors contributed equally to this work.
†Present address: Max-Planck-Institut für Medizinische Forschung, Abteilung für Biomolekulare Mechanismen, Jahnstrasse 29, D-69120 Heidelberg, Germany.
‡To whom correspondence should be addressed. E-mail: messinger@mpi-muelheim.mpg.de (J.M.); zouni@phosis1.chem.tu-berlin.de (A.Z.); vkyachandra@lbl.gov (V.K.Y.)

troscopy (*11*). In addition, recent x-ray diffraction (XRD) studies of single crystals of PSII provide critical information about its structure at 3.8 to 3.0 Å resolution (*12*–*16*). However, even XRD data of the highest resolution presently available are insufficient to accurately determine the positions of Mn, Ca, and the bridging and terminal ligands. This is reflected by the differences in the placement of the metal ions and putative ligands in the 3.0 (*16*) and 3.5 Å (*14*) structures. Furthermore, at the x-ray dose and temperature used in the XRD studies, the geometry of the Mn_4Ca cluster is disrupted, initiated by the rapid reduction of Mn(III) and Mn(IV) present in the dark-stable S_1 state to Mn(II), as shown by Mn x-ray absorption near-edge structure (XANES) studies and Mn x-ray absorption fine structure (EXAFS) studies of PSII single crystals (*17*).

EXAFS experiments with PSII require a substantially lower x-ray dose than XRD measurements (*17*), and the onset of radiation damage can be precisely determined and controlled by monitoring the Mn K-edge position, thus allowing the collection of data from the intact Mn_4Ca cluster of PSII. In addition, EXAFS provides metal-to-metal and metal-to-ligand distances with high accuracy (~0.02 Å) and a resolution of ~0.1 Å. Mn and Ca EXAFS studies of frozen solutions of PSII preparations have provided accurate distances and numbers of Mn-Mn, Mn-Ca, and Mn/Ca-ligand vectors in the Mn_4Ca cluster (*10*, *18*–*21*) and have led to the development of several possible structural models for the Mn_4Ca cluster (fig. S1).

For polarized EXAFS experiments on one-dimensionally oriented membranes or three-dimensionally oriented single crystals, the EXAFS amplitude is orientation dependent and proportional to $\sim\cos^2\theta$, where θ is the angle between the **e**-field vector of the polarized x-ray beam and the absorber-backscatterer vector. Therefore, this technique provides important additional geometric information about the metal site in metalloprotein crystals (*22*, *23*). Here we show that polarized EXAFS can be used to provide structural models of the Mn_4Ca active site, revealing details currently unresolvable by XRD. Combining information from polarized EXAFS and XRD (*16*) leads to the placement of these models within the PSII protein environment.

PSII crystals from *Thermosynechococcus elongatus* in the dark-stable S_1 state were oriented so that the x-ray **e**-field vector was parallel to the *a*, *b*, or *c* axis of the crystal unit cell. The total photon dose on the sample was kept at a safe level of 1×10^7 photons/μm^2 on the basis of detailed radiation damage studies on single crystals of PSII (*17*). After the Mn XANES and EXAFS spectra were collected, the orientation of the crystal was determined in situ by the collection of x-ray diffraction patterns (fig. S2) (*24*).

Figure 1A (top) shows the polarized Mn K-edge XANES spectra of PSII single crystals with the x-ray **e**-field vector parallel to each of the orthogonal crystal unit cell axes (*a*, *b*, and *c*). The spectra show distinctive features at each orientation, both in the main K edge (1s to 4p) and the pre-edge region [1s-to-3d transitions (Fig. 1A, inset)]. The orientation-dependent differences are more clearly seen in the second-derivative plots (Fig. 1A, bottom).

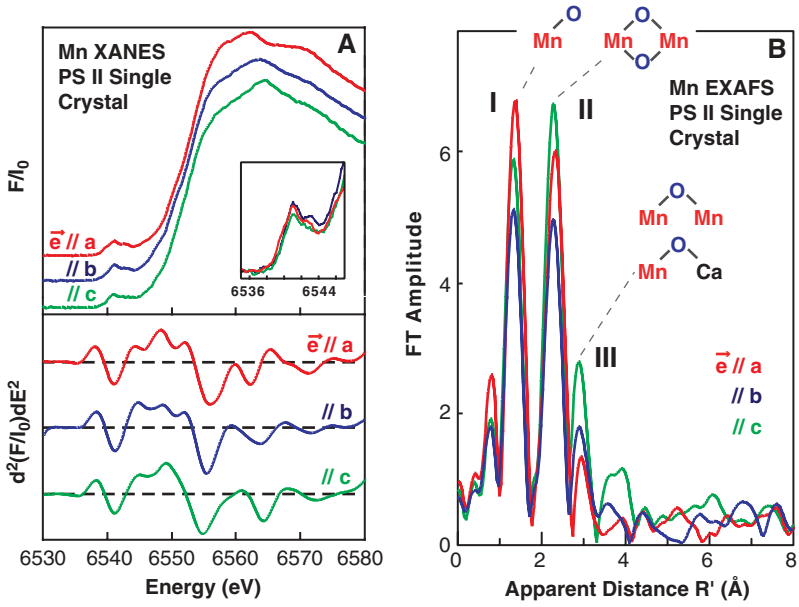
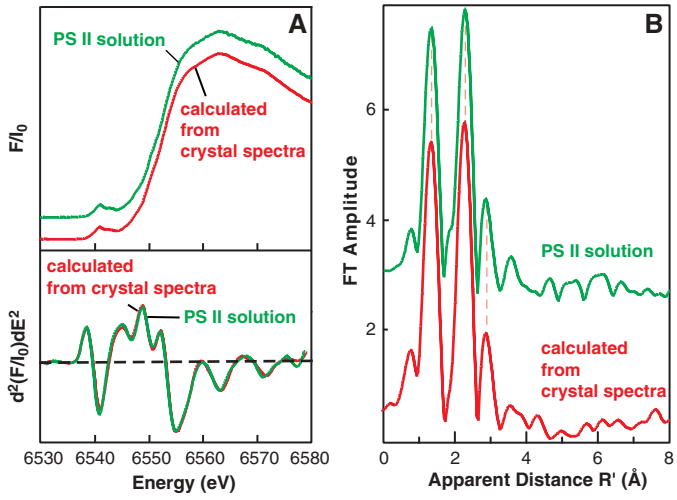


Fig. 1. (**A**) Single-crystal, polarized Mn K-edge XANES spectra (top) and the corresponding second derivatives of the XANES spectra (bottom) of PSII in the S_1 state. The XANES spectra are shown with the x-ray **e**-field vector aligned parallel to the principal crystal axes from single crystals of PSII [*a* (red curve), *b* (blue curve), and *c* (green curve)] from *T. elongatus*. The inflection point of the edges and the shape of the spectra are clearly dependent on the orientation of the single crystal with respect to the x-ray **e** vector. The inset shows the dichroism of the pre-edge region assigned to the 1s-to-3d transition. F/I_0, absorbance; *E*, energy. (**B**) FTs of polarized Mn EXAFS spectra from single crystals of PSII in the S_1 state. The FTs are from EXAFS spectra with the x-ray **e**-field vector aligned parallel to the crystal unit cell axes of PSII [*a* (red curve), *b* (blue curve), and *c* (green curve)]. Each of the three FT peaks characteristic of Mn EXAFS from PSII is dichroic. FT peak I is from Mn-ligand backscattering; FT peak II is from three Mn-Mn distances at 2.7 to 2.8 Å; and FT peak III is from one Mn-Mn and two Mn-Ca distances at 3.3 and 3.4 Å, respectively. All Fourier peaks appear at an apparent distance R′ that is shorter than the actual distance R by ~0.5 Å due to a phase shift. The dichroism of the metal-to-metal distances reflects the geometry of the Mn_4Ca cluster.

Fig. 2. (**A**) Comparison of Mn XANES spectra from solution PSII samples in the S_1 state (*T. elongatus*) (green curve) with the powder spectrum (red curve) calculated from the three PSII single-crystal, polarized spectra (Fig. 1A). The corresponding second derivatives of the Mn XANES (bottom) reveal that the XANES spectra of the solution and the calculated powder spectra from polarized PSII single-crystal spectra are indistinguishable. (**B**) Comparison of the FTs of the Mn EXAFS spectrum from solution PSII samples (green curve) and the powder spectrum (red curve) calculated from the three PSII single-crystal, polarized EXAFS spectra (Fig. 1B). The FTs are indistinguishable within the signal-to-noise ratio.



The Fourier transforms (FTs) of the polarized Mn EXAFS of PSII single crystals are shown in Fig. 1B. Peak I has contributions from bridging and terminal O or N atoms at 1.8 to 2.0 Å; peak II results from di-μ-oxo–bridged Mn-Mn interactions at 2.7 to 2.8 Å; and peak III contains contributions from mono-μ-oxo–bridged Mn-Mn and Mn-Ca interactions at 3.3 and 3.4 Å, respectively. The EXAFS spectra show a pronounced dependence on the crystal orientation, indicating that the Mn_4Ca cluster is highly asymmetric.

A test of whether the redox state and the structure of the Mn_4Ca cluster in the single crystals are equivalent to that in active oxygen-evolving PSII solution samples is to calculate the isotropic solution spectra (powder spectra) from the single-crystal spectra and compare these to experimental solution spectra. The comparison of XANES and EXAFS spectra in Fig. 2, A and B (with the corresponding second derivatives for XANES in Fig. 2A, bottom), shows that the spectra are indistinguishable within the extremely low noise of these experimental data. This result confirms the dichroism of the EXAFS spectra and shows that the cluster geometry is not disrupted by radiation damage. Therefore, the data of Fig. 1B form a reliable basis for the elucidation of the structure and orientation of the Mn_4Ca cluster within PSII.

Initially, we used the polarized EXAFS data to test the Mn_4Ca structures proposed on the basis of XRD data (*14*, *16*). The comparison of the experimental data (Fig. 1B) with the polarized EXAFS spectra calculated for the XRD models at 3.5 (*14*) and 3.0 Å (*16*) resolution is shown in fig. S3. The disagreement is notable; it is most likely due to the limited resolution and x-ray damage to the Mn_4Ca cluster during XRD measurements (*17*).

In the next step, the polarized EXAFS data were used to evaluate and filter the large number of previously proposed and feasible models for the Mn_4Ca cluster [Mn_4 motifs **a** to **r** (fig. S1)], including topologically related structures [the motifs from the XRD structures are **c** (*16*) and **n** (*14*)]. For this filtering process, the polarized EXAFS for the proposed structural models was calculated as follows (fig. S4) (*24*): (i) Each model was placed into one PSII monomer, (ii) the coordinates for the companion monomer were determined by means of the local C_2 symmetry of the dimeric PSII complex, (iii) the coordinates for the four symmetry-related dimeric units within the $P2_12_12_1$ crystal unit cell were determined, and (iv) the EXAFS spectrum was calculated for the PSII single crystal with the **e** vector of the x-ray beam parallel to each of the crystal unit cell axes (*a*, *b*, and *c*) with the use of the FEFF8 program (*25*). To determine the optimal orientation, we rotated each of the inserted clusters in a stepwise fashion within the PSII monomer with respect to the three axes *x*, *y*, and *z* [for definition of axes, see Fig. 3 (top right)]. For each orientation, the EXAFS spectra were calculated and compared to the experimental results shown in Fig. 1B. This process was continued until the rotational space was mapped sufficiently to determine whether a specific model complied with the polarized EXAFS data (*24*). The focus in our comparison was predominantly on the contribution to the EXAFS from Fourier peaks II and III, which are from Mn-Mn and Mn-Ca interactions that mainly define the motif for the Mn_4Ca cluster. This process rules out unsatisfactory cluster geometries with a high degree of confidence.

Range-extended EXAFS results (*21*) show that three short Mn-Mn distances between 2.7 and 2.8 Å are present in the Mn_4Ca cluster. The presence of the 2.8 Å Mn-Mn vector is observed as a shift of FT peak II to a longer distance along the *a* axis in the polarized EXAFS data (Fig. 1B). In addition to the 2.8 Å vector, two 2.7 Å vectors are required to reproduce the intensity of the FT peak II, mainly along the *b* and *c* axes (see the EXAFS fit parameters in table S1). Therefore, the models containing only two short 2.7 Å Mn-Mn distances, or one 2.7 Å and one 2.8 Å Mn-Mn distances [motifs **a** to **j** (fig. S1A)], were unacceptable in reproducing both the solution range-extended EXAFS (*21*) and polarized EXAFS data.

The remaining models contain three Mn-Mn vectors at 2.7 to 2.8 Å and 0, 1, or 2 Mn-Mn vectors at 3.3 Å [motifs **k** to **r** (fig. S1B)]. The Mn EXAFS solution of isotropic frozen PSII

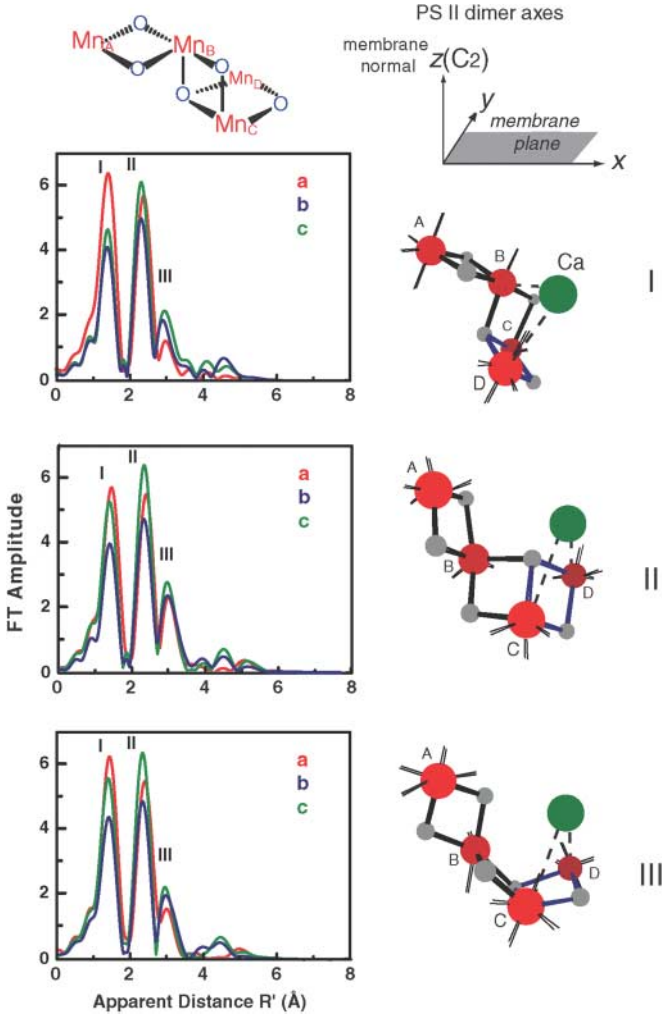


Fig. 3. High-resolution structural models for the Mn_4Ca cluster in PSII from polarized EXAFS. The Mn_4Ca models I, II, and III are shown on the right. Each model is compatible with the polarized Mn EXAFS spectra of single crystals of PSII. The Mn_4 motif (**r** in fig. S1) common to the three structures is shown at upper left. The models are shown in the orientation in which they should be placed in the PSII membrane according to the axis system shown at upper right. The *z* axis is chosen to be along the membrane normal. The *x* and *y* axes define the membrane plane; the *x* axis connects the two nonheme Fe atoms of the dimer that are related by the noncrystallographic local C_2 axis. The particular orientations shown for models I, II, and III were chosen on the basis of their compatibility with the overall electron density and the positioning of the protein ligands in the 3.0 Å resolution x-ray crystal structure (*16*). The Mn atoms are shown in red. The distance between Mn_C and Mn_D atoms is ~2.8 Å (indicated by blue oxo bonds), and the distance between the Mn_A and Mn_B atoms, as well as the Mn_B and Mn_C atoms, is ~2.7 Å. The distance between Mn_B and Mn_D is ~3.3 Å. The Ca atom (green sphere) is ~3.4 Å from two Mn atoms [in model II, Ca can be at ~3.4 Å to Mn_D and Mn_C or Mn_D and Mn_B (model IIa, fig. S5)]. The bridging motif to Ca is not well defined by our experiments; therefore, dashed lines connect the Ca atom to the two Mn atoms at ~3.4 Å. The experimental FTs from the polarized Mn EXAFS spectra from single crystals (Fig. 1B) are comparable to the FTs calculated for these models. The dichroism of the EXAFS spectra will be identical for local C_2-related orientations.

solution samples shows that FT peak III is fit best by one Mn-Mn vector at 3.3 Å and two Mn-Ca vectors at 3.4 Å (*10*), which is supported by Sr and Ca EXAFS results (*19*, *20*). Models with these distances [motifs **p** to **r** (fig. S1B)] and models containing two 3.3 Å Mn-Mn distances [motifs **l** to **o** (fig. S1B)] along with the Mn-Ca distances were included in our search for a match with the observed dichroism (a heterogeneity of ~0.2 Å was included for models with two 3.3 Å Mn-Mn vectors to decrease the otherwise too-intense contribution to FT peak III).

The application of the rigorous testing protocol described above showed that the polarized EXAFS data (Fig. 1B) are consistent with three topologically related structures [models I, II, and III (Fig. 3)] that are based on the Mn_4 motif **r** (Fig. 3, top) (*18*, *21*). These models are shown in the orientation (indicated by the axis system shown in Fig. 3, top right) required to satisfy the polarized EXAFS data. The experimental polarized spectra (Fig. 1B) and the calculated spectra from the three structures positioned as shown (Fig. 3, left panel) are very similar with regard to the intensity and the orientation dependence for FT peak II and III, which determine the motif of the structural model. The trend in the dichroism is also similar for FT peak I. Most of the contribution to peak I is from bridging O atoms, because Mn-terminal ligand distances are highly disordered (*10*, *18*) (table S1) and hence were not included in the simulations. The ambiguities in the intensities in FT peak I, however, come from the small contributions from this disordered shell of terminal ligands, which are difficult to model.

At present, it is not possible to choose among the three models on the basis of the polarized EXAFS data. In addition, each of these structural models can have four symmetry-related orientations in the membrane originating from the $\cos^2\theta$ dependence of the EXAFS signal and the noncrystallographic C_2 symmetry of the monomers in the PSII dimer (*24*).

To further distinguish between these structural options, we used the following criteria: (i) the placement of the Ca atoms; (ii) the distribution of the electron density from XRD (*16*), which is pear-shaped, with the narrow end oriented toward the dimer C_2 axis; and (iii) the information from the position of putative amino acid ligands identified around the electron density coinciding with the Mn_4Ca cluster (*16*). The placement of Ca relied on Sr EXAFS data showing that the two Mn-Sr(Ca) vectors are along the membrane normal (*20*) and the anomalous XRD studies indicating that Ca is above the Mn plane toward the redox-active tyrosine Y_Z and P_{680} chlorophylls (*14*, *16*). The placement of Ca reduces the options among the symmetry-related orientations by half. The remaining two symmetry-equivalent orientations of models I, II, and III were examined for compliance with the overall trend of the electron density and the placement of ligands (*24*). The best orientations of the three models within the PSII membrane are shown in Fig. 3. The other symmetry-related orientation is clearly unsatisfactory because of the pronounced asymmetry in the electron density.

The best-fit ligand environment was obtained for all the models [coordinates in table S2 (*24*)], but there was some overlap between the cluster and ligand positions (fig. S6). The ligand assignments must remain tentative at this point because movement of the protein ligand residues can result from radiation damage to the metal site during XRD (*17*). In the XRD study, the Mn close to Asp^{170} (i.e., Mn_A) and Ca have a lower occupancy (σ value) as compared to those of the other Mn atoms (*16*). This lower value might indicate that the Mn_A and Ca metal sites are more susceptible to radiation damage and hence their ligands may be prone to movement. The ligand environment for the best-fit orientation for model II is shown in Fig. 4.

The considerable differences in the orientation and structures between model II and those in the recent XRD proposals (*14*, *16*) for the Mn_4Ca complex are shown as an overlay in fig. S7 (*24*). This highlights the difficulties of deriving high-resolution structures from low-resolution XRD. Although constraints from EXAFS distance information are included in the building of the XRD model that lead to approximate structures, there are too many structures that can be fit to the low-resolution electron density. The importance of polarized EXAFS of single crystals lies in its ability to

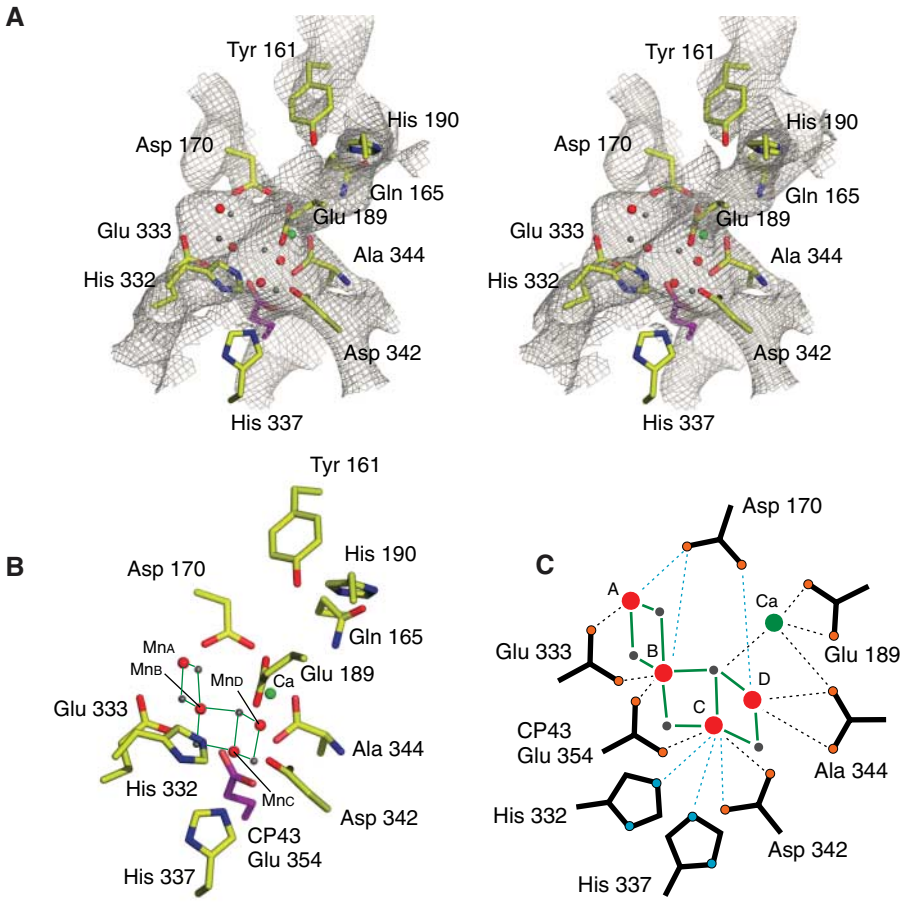


Fig. 4. Placement of model II for the Mn_4Ca cluster derived from polarized Mn EXAFS in relation to the putative ligands obtained from the 3.0 Å resolution x-ray crystal structure (*16*). The spheres represent Mn (red), Ca (green), and the bridging oxygen ligand atoms (gray). The assignment of ligands is tentative because it is based on the electron density of the Mn_4Ca cluster, and its immediate environment may be altered by x-ray damage. (**A**) Stereo view of the electron density (*16*) with model II placed in the density, as explained in the text. The view is along the membrane plane. (**B**) Same view as in (A) with ligands and atoms labeled correspondingly. All ligands are from D1 except for CP43-Glu^{354}. Electron density was omitted for clarity. (**C**) A schematic diagram of model II with the view along the membrane plane, with putative ligands from the electron density. Bonds between Mn and the bridging oxo are shown as solid green lines. Bonds to putative terminal ligand atoms from amino acids and to Ca atoms are shown as dotted lines (black, less than 3.0 Å; blue, more than 3.0 Å).

discriminate between the many possible high-resolution models by relying on the dichroism of the EXAFS spectra.

The structural changes of the Mn_4Ca complex on advancing through the S_i state intermediates can be placed in the context of the polarized EXAFS data to assist in deriving a mechanism for photosynthetic water oxidation. The FTIR data, in conjunction with model II (Fig. 4), suggest that Mn_A, which may be ligated by Asp^{170}, does not change oxidation state and remains Mn(III) or Mn(IV) throughout the Kok cycle. The C-terminal Ala^{344} may be a ligand to Mn_D, which is proposed to undergo Mn(III)→Mn(IV) oxidation during the $S_1 \rightarrow S_2$ transition (*26*–*28*). Recent FTIR data suggest that His^{332} monitors structural changes of the Mn_4Ca cluster, but no evidence for a Mn-centered oxidation was reported (*29*). Because Mn_C is closer to His^{332}, Mn_C may remain Mn(III) or Mn(IV) throughout the cycle. Consequently, Mn_B is a likely candidate for Mn oxidation during the $S_0 \rightarrow S_1$ transition.

The dichroism in the polarized EXAFS data from single crystals provides a powerful filter for choosing among many of the proposed structural models. Also, as shown in this study, the combination of XRD and polarized EXAFS on single crystals has several advantages for unraveling structures of x-ray damage–prone, redox-active metal sites in proteins. XRD structures at medium resolution are sufficient to determine the overall shape and placement of the metal site within the ligand sphere, and refinement by means of polarized EXAFS can provide accurate metal-to-metal and metal-to-ligand vectors. In addition, different intermediate states of the active site (including different metal oxidation states), which may be difficult to study with XRD at high resolution, can be examined. The structural model from polarized EXAFS from the S_1 state presented here, and from the other S states, will provide a reliable foundation for the investigation of the mechanism of photosynthetic water oxidation and for the design of biomimetic catalysts for water splitting.

References and Notes

1. T. Wydrzynski, S. Satoh, *Photosystem II: The Light-Driven Water: Plastoquinone Oxidoreductase, Advances in Photosynthesis and Respiration Series,* vol. 22 (Springer, Dordrecht, Netherlands, 2005).
2. B. Kok, B. Forbush, M. McGloin, *Photochem. Photobiol.* **11**, 457 (1970).
3. T. G. Carrell, A. M. Tyryshkin, G. C. Dismukes, *J. Biol. Inorg. Chem.* **7**, 2 (2002).
4. G. W. Brudvig, *Adv. Chem. Ser.* **246**, 249 (1995).
5. K. Hasegawa, T.-A. Ono, Y. Inoue, M. Kusunoki, *Chem. Phys. Lett.* **300**, 9 (1999).
6. J. M. Peloquin, R. D. Britt, *Biochim. Biophys. Acta* **1503**, 96 (2001).
7. J. Messinger, J. H. A. Nugent, M. C. W. Evans, *Biochemistry* **36**, 11055 (1997).
8. K. A. Åhrling, S. Peterson, S. Styring, *Biochemistry* **37**, 8115 (1998).
9. L. V. Kulik, B. Epel, W. Lubitz, J. Messinger, *J. Am. Chem. Soc.* **127**, 2392 (2005).
10. V. K. Yachandra, K. Sauer, M. P. Klein, *Chem. Rev.* **96**, 2927 (1996).
11. H.-A. Chu, W. Hillier, N. A. Law, G. T. Babcock, *Biochim. Biophys. Acta* **1503**, 69 (2001).
12. A. Zouni *et al.*, *Nature* **409**, 739 (2001).
13. N. Kamiya, J. R. Shen, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 98 (2003).
14. K. N. Ferreira, T. M. Iverson, K. Maghlaoui, J. Barber, S. Iwata, *Science* **303**, 1831 (2004).
15. J. Biesiadka, B. Loll, J. Kern, K. D. Irrgang, A. Zouni, *Phys. Chem. Chem. Phys.* **6**, 4733 (2004).
16. B. Loll, J. Kern, W. Saenger, A. Zouni, J. Biesiadka, *Nature* **438**, 1040 (2005).
17. J. Yano *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12047 (2005).
18. J. H. Robblee *et al.*, *J. Am. Chem. Soc.* **124**, 7459 (2002).
19. R. M. Cinco *et al.*, *Biochemistry* **41**, 12928 (2002).
20. R. M. Cinco *et al.*, *Biochemistry* **43**, 13271 (2004).
21. J. Yano *et al.*, *J. Am. Chem. Soc.* **127**, 14974 (2005).
22. R. A. Scott, J. E. Hahn, S. Doniach, H. C. Freeman, K. O. Hodgson, *J. Am. Chem. Soc.* **104**, 5364 (1982).
23. A. M. Flank, M. Weininger, L. E. Mortenson, S. P. Cramer, *J. Am. Chem. Soc.* **108**, 1049 (1986).
24. Materials and methods are available as supporting material on *Science* Online.
25. J. J. Rehr, R. C. Albers, *Rev. Mod. Phys.* **72**, 621 (2000).
26. H. A. Chu, W. Hillier, R. J. Debus, *Biochemistry* **43**, 3152 (2004).
27. R. J. Debus, M. A. Strickler, L. M. Walker, W. Hillier, *Biochemistry* **44**, 1367 (2005).
28. Y. Kimura, N. Mizusawa, T. Yamanari, A. Ishii, T. Ono, *J. Biol. Chem.* **280**, 2078 (2005).
29. Y. Kimura, N. Mizusawa, A. Ishii, T. Ono, *Biochemistry* **44**, 16072 (2005).

30. This work was supported by the Director, the Office of Science, the Office of Basic Energy Sciences (OBES), and the Division of Chemical Sciences, Geosciences, and Biosciences of the U.S. Department of Energy (DOE) under contract DE-AC02-05CH11231; by NIH grant GM 55302; by the Deutsche Forschungsgemeinschaft (grants SFB 498, TP A4, C7, and Me 1629/2–3); and by the Max- Planck-Gesellschaft. Synchrotron facilities were provided by the Stanford Synchrotron Radiation Laboratory (SSRL) operated by DOE OBES. The SSRL Biomedical Technology Program is supported by NIH, the National Center for Research Resources, and the DOE Office of Biological and Environmental Research. We thank E. Berry (Lawrence Berkeley National Laboratory) for help with indexing the PSII XRD pattern and many insights; D. DiFiore and C. Lüneberg for technical assistance; and U. Bergmann (SSRL), S. Un (Commissariat à l'Energie Atomique), and P. Glatzel (European Synchrotron Radiation Facility) for discussions about the analysis of polarization data. We are grateful to M. P. Klein (1921–2000) and H.-T. Witt for their encouragement in initiating this collaborative research.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/821/DC1
Material and Methods
SOM Text
Figs. S1 to S7
Tables S1 and S2
References

31 March 2006; accepted 14 June 2006
10.1126/science.1128186

Transgenic Mice with a Reduced Core Body Temperature Have an Increased Life Span

Bruno Conti,[1,2]* Manuel Sanchez-Alavez,[1,2] Raphaelle Winsky-Sommerer,[3]† Maria Concetta Morale,[1]‡ Jacinta Lucero,[1,2] Sara Brownell,[1,2]§ Veronique Fabre,[3]‖ Salvador Huitron-Resendiz,[2] Steven Henriksen,[2]¶ Eric P. Zorrilla,[1,2] Luis de Lecea,[3]# Tamas Bartfai[1,2]

Reduction of core body temperature has been proposed to contribute to the increased life span and the antiaging effects conferred by calorie restriction (CR). Validation of this hypothesis has been difficult in homeotherms, primarily due to a lack of experimental models. We report that transgenic mice engineered to overexpress the uncoupling protein 2 in hypocretin neurons (Hcrt-UCP2) have elevated hypothalamic temperature. The effects of local temperature elevation on the central thermostat resulted in a 0.3° to 0.5°C reduction of the core body temperature. Fed ad libitum, Hcrt-UCP2 transgenic mice had the same caloric intake as their wild-type littermates but had increased energy efficiency and a greater median life span (12% increase in males; 20% increase in females). Thus, modest, sustained reduction of core body temperature prolonged life span independent of altered diet or CR.

Temperature homeostasis in mammals is regulated centrally by neurons located in the preoptic area (POA) of the hypothalamus, a region that includes the medial and lateral part of the preoptic nucleus, the anterior hypothalamus, and the nearby regions of the septum. This region is believed to contain the central thermostat, which keeps core body temperature (CBT) within a very narrow range even when the animal is exposed to a wide range of ambient temperatures. Lesion and thermal stimulation studies have demonstrated that the POA senses changes in local and peripheral temperatures and coordinates thermoregulatory responses [for review, see (*1*)].

With the aim of generating animals with a reduced CBT, we hypothesized that local heat production within or proximate to the POA, by mimicking an increase in CBT, might ac-

tivate compensatory thermoregulatory mechanisms and thereby reduce CBT. We generated mice that overexpressed the uncoupling protein 2 (UCP2) exclusively in hypocretin (Hcrt) neurons (Hcrt-UCP2 mice). UCP2 is an inner mitochondrial membrane protein that uncouples oxidative phosphorylation from respiration by leaking hydrogen ions from the intermembrane space to the matrix, thereby dissipating the proton gradient energy in the form of heat (*2*). Hypocretins (hypocretin 1 and 2), also known as orexins, are neuropeptides derived from a common precursor that participate in the regulation of the sleep/wake cycle, energy balance, food intake, and endocrine and autonomic functions (*3*–*5*). Hypocretins are exclusively expressed in approximately 3000 neurons in the lateral hypothalamus (LH) at a distance of 0.8 mm from the POA (*6*) and provide an anatomically restricted site for heat generation in the hypothalamus.

A colony of Hcrt-UCP2 mice and wild-type littermates was established from one founder generated on and backcrossed for seven generations on a C57/BL6 background (fig. S1A).

To determine whether UCP2 overexpression in hypocretin neurons resulted in heat production that could affect the POA, we measured local temperature in the LH and in the POA. Local temperature was significantly higher in both the LH and the POA of Hcrt-UCP2 mice compared with wild-type mice (Fig. 1). Temperature elevation averaged 0.65°C in the LH and 0.32°C in the POA. The smaller increase in temperature elevation in the POA compared with the LH was likely due to heat dissipation from the LH. The difference between LH and POA temperature was constant during the 24 hours of recording.

The effect of elevated hypothalamic temperature on CBT was studied using radiotelemetry in male and female mice (Fig. 2). Hcrt-UCP2 mice maintained a normal circadian variation of CBT during the light and the dark cycles. In males, no difference in the CBT values between Hcrt-UCP2 and wild-type mice was observed during the light phase or during the transition between phases. However, Hcrt-UCP2 mice consistently exhibited a significantly lower CBT during the dark phase throughout several days of recording (Fig. 2). In females, the reduction of CBT averaged 0.34°C and was more pronounced in the second part of the dark phase with a peak difference of 0.6°C (Fig. 2, A and B). A similar pattern was seen in male transgenics, with no difference observed during the light phase and the transition from light to dark, but an average CBT reduction of 0.3°C and a peak difference of 0.56°C observed during the dark phase (Fig. 2, D and E). In contrast to males, female Hcrt-UCP2 mice also showed a significant reduction of CBT in the first half of the transition from dark to light. Motor activity was similar between Hcrt-UCP2 and wild-type mice, being only marginally higher in Hcrt-UCP2 male mice at the end of the light-dark transition, a time when no difference in CBT was observed. In females, motor activity of Hcrt-UCP2 mice was marginally lower in the last part of the dark phase, when CBT also was lowest (Fig. 2, C and F). After injection with *Escherichia coli* lipopolysaccharides (LPS), Hcrt-UCP2 mice developed a fever response similar in amplitude and duration to that of the wild-type mice, which indicates that the thermogenic capacity of Hcrt-UCP2 mice was not impaired (fig. S3). The CBT profile was identical between transgenic and wild-type mice during the stress peak and the light phase, but Hcrt-UCP2 mice

[1]Harold L. Dorris Neurological Research Center, Scripps Research Institute, La Jolla, CA 92037, USA. [2]Molecular and Integrative Neurosciences Department, Scripps Research Institute, La Jolla, CA 92037, USA. [3]Department of Molecular Biology, Scripps Research Institute, La Jolla, CA 92037, USA.

*To whom correspondence should be addressed. E mail: bconti@scripps.edu

†Present address: Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland.

‡Present address: Dipartimento di Neurofarmacologia, OASI (IRCCS), Troina, Italy.

§Division of Geriatric Medicine, Department of Medicine, UCSD, La Jolla, CA 92037, USA.

‖Present address: Unité Mixte de Recherche 677, Institut National de la Santé et de la Recherche Médicale/Universite Pierre et Marie Curie, Paris, France.

¶Present address: Western University of Health Sciences, Pomona, CA 91766, USA.

#Present address: Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, CA 94305, USA.

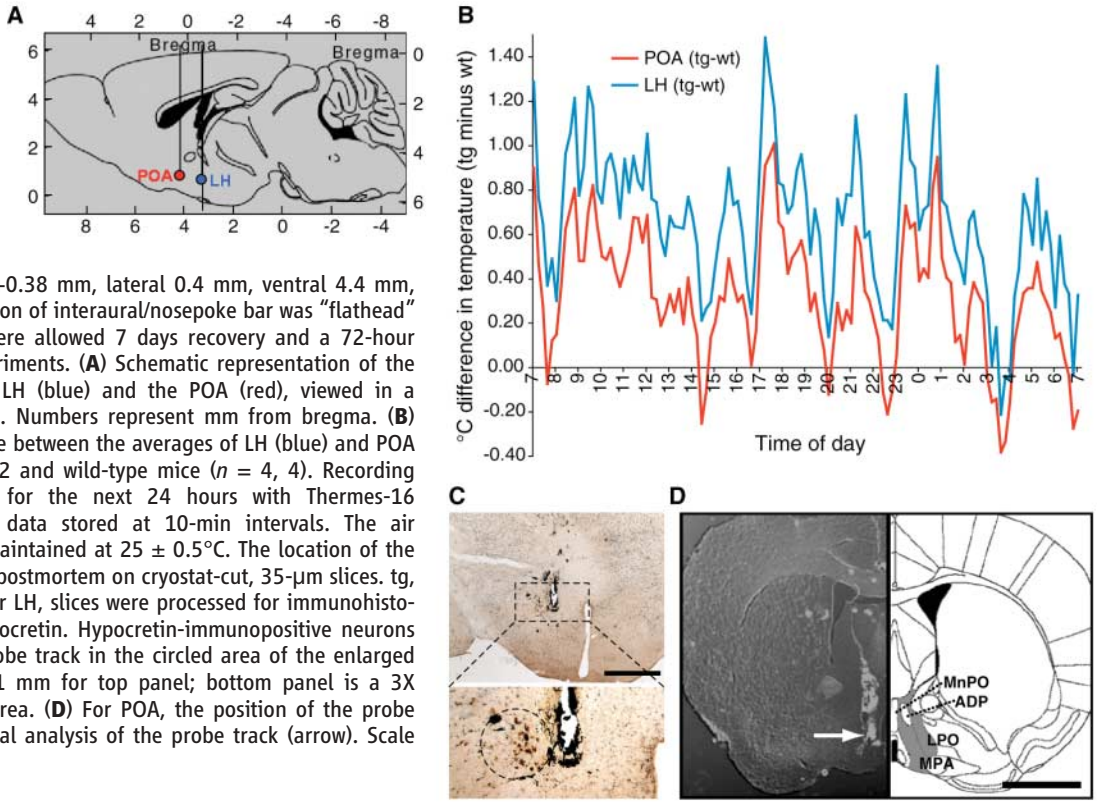


Fig. 1. UCP2 overexpression elevated local temperature. Three-month-old mice were stereotactically implanted with thermocouple probes in the lateral hypothalamic area (anteroposterior from bregma –0.7 mm, lateral 1.25 mm, ventral 4.5 mm, dura at point of entry) and the POA (anteroposterior from bregma +0.38 mm, lateral 0.4 mm, ventral 4.4 mm, dura at point of entry; the position of interaural/nosepoke bar was "flathead" bregma = lambda). Animals were allowed 7 days recovery and a 72-hour habituation period before experiments. (**A**) Schematic representation of the location of the probes in the LH (blue) and the POA (red), viewed in a sagittal section of mouse brain. Numbers represent mm from bregma. (**B**) Profile of temperature difference between the averages of LH (blue) and POA (red) temperatures of Hcrt-UCP2 and wild-type mice (n = 4, 4). Recording was carried out continuously for the next 24 hours with Thermes-16 (Physitemp Instruments), with data stored at 10-min intervals. The air temperature in the room was maintained at 25 ± 0.5°C. The location of the recording sites was determined postmortem on cryostat-cut, 35-μm slices. tg, transgenic; wt, wild-type. (**C**) For LH, slices were processed for immunohistochemistry with antibody to hypocretin. Hypocretin-immunopositive neurons are visible at the left of the probe track in the circled area of the enlarged detailed section. Scale bar is 1 mm for top panel; bottom panel is a 3X magnification of the selected area. (**D**) For POA, the position of the probe was demonstrated by histological analysis of the probe track (arrow). Scale bar, 2 mm.

maintained a temperature slightly higher than wild-type mice during the first half and the end of the dark phase. Overall, the data indicate that the reduction of basal CBT observed in Hcrt-UCP2 mice did not result from reduced locomotor activity or impaired thermogenic ability, but is consistent with an effect on the central thermostat.

We found that UCP2 overexpression reduced the number of hypocretin immunoreactive neurons by 22% and 30% in male and female Hcrt-UCP2 mice, respectively (fig. S4). It might be argued that intracellular temperature elevation, an excessive reduction in ATP synthesis, or altered intracellular Ca^{2+} concentrations resulting from UCP2 overexpression interferes with the normal metabolic activity of hypocretin neurons. Intracerebroventricular injection of pharmacologically high doses of hypocretin 1 reportedly elevates spontaneous physical activity and CBT (*7*–*10*). However, the possibility that a decreased number of hypocretin neurons contributed to the reduced CBT was ruled out in orexin/ataxin-3 mice (*11*) that showed 90% reduction of hypocretin neurons but no significantly lowered CBT (fig. S5). No differences in sleep parameters that could account for the reduction of CBT were found (SOM text).

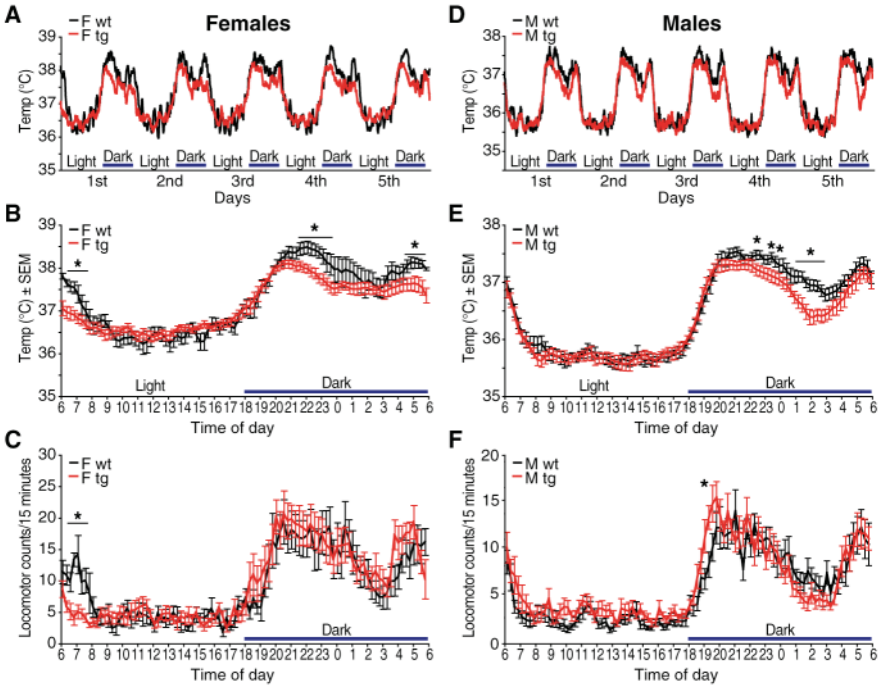


Fig. 2. Core body temperature and motor activity profile. CBT and motor activity were recorded simultaneously on mice implanted intraperitoneally with radiotelemetric transmitters (Data Science International, St. Paul, MN). Mice were allowed to recover for 2 weeks after surgery before recording. The recording was carried out over a period of 10 days. Panels show 5-day temperature profiles (**A** and **D**), the average temperature (**B** and **E**), and locomotor activity profiles (**C** and **F**) from 10 days of 24-hour recording in both male and female mice. The air temperature of the room was maintained at 25 ± 0.5°C [n = 15 mice per group for males, n = 5 mice per group for females, *P < 0.05 analysis of variance (ANOVA) with repeated measures followed by Fisher's least significant difference (LSD) tests comparing 10-min intervals]. tg, transgenic; wt, wild type.

The effects of UCP2 overexpression in hypocretin neurons on water and food consumption were also measured. Hcrt-UCP2 mice did not differ from wild-type mice in their intake of chow (measured every 3 hours or biweekly) (Fig. 3A) or water (M ± SEM: 3.7 ± 0.3 versus 3.6 ± 0.2 ml for wild-type and Hcrt-UCP2, measured biweekly). Whereas body weights of female Hcrt-UCP2 and wild-type mice did not significantly differ, male transgenic mice began to weigh significantly more than wild-type mice beginning at 20 weeks of age. By 35 weeks of age, the male transgenics weighed 10% more than the wild-type males (Fig. 3B). When subjected to 27 hours of food deprivation, Hcrt-UCP2 transgenic mice lost significantly less weight than would be predicted from their expected metabolic body mass demands as compared with wild-type littermates (genotype effect: $F(1,29) = 20.64$, $P < 0.0001$). The decrease in putative relative energy expenditure, which was similar in both male and female transgenics (Fig. 3C), is an index of increased metabolic efficiency most likely reflecting the reduced energy required to maintain a lower CBT (*12*).

Reduction of CBT has antiaging effects and prolongs life in poikilotherms (*13*). In homeotherms, reduction of CBT results from calorie restriction (CR), a controlled dietary regimen that prolongs life span in rodents (*14*, *15*) and that has been reported to delay the onset of a variety of diseases in model organisms (*16*–*21*). However, whether re-

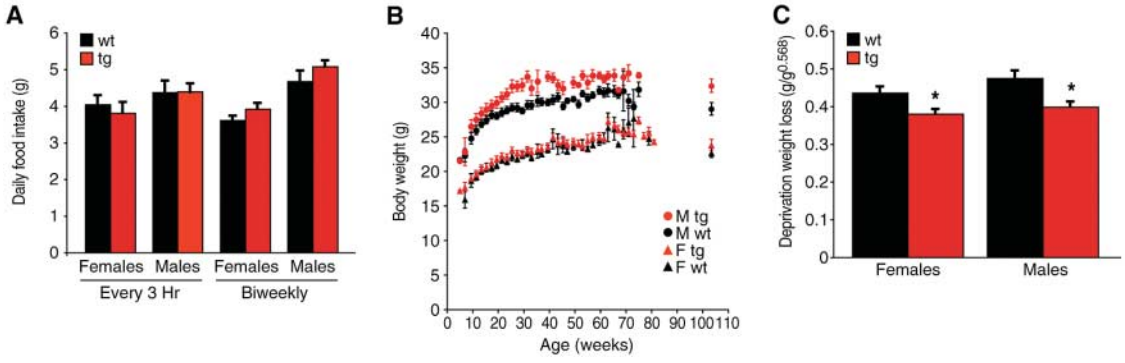


Fig. 3. Food intake, growth curve, and energy expenditure. (**A**) Mean (±SEM) daily chow (11% kcal fat) intake of mice as estimated from measuring food every 3 hours over a 24-hour period or biweekly over a 2-week period (as indicated) in adult age-matched (5 to 6 months old) and weight-matched Hcrt-UCP2 mice (n = 9 males and 9 females). Mice from the two genotypes did not differ in their food intake. (**B**) Growth curve of male and female Hcrt-UCP2 and wild-type littermates determined from biweekly measurements of body weight (n = 9 mice per group; *P < 0.05 ANOVA). (**C**) Mean (±SEM) weight loss after 27 hours of food deprivation in Hcrt-UCP2 and wild-type control mice, normalized for the expected metabolic demands of their predeprivation body mass as estimated by a power function (g weight loss/g baseline weight$^{0.568}$) (*29*). Male and female transgenic mice lost 13 and 16% less weight, respectively, during food deprivation than would have been expected from their metabolic body mass as determined by weight loss of wild-type controls (n = 6 and 9, *P < 0.005, Fisher's protected LSD test). tg, transgenic; wt, wild-type.

duced CBT in itself prolongs life span in homeotherms has not been demonstrated. To investigate this question, we compared the survivorship of Hcrt-UCP2 mice with wild-type littermates fed ad libitum on an 11% fat (kcal) diet. Despite eating normally (Fig. 3A), the Hcrt-UCP2 genotype showed a 25% reduction in mortality rate across adulthood. As a consequence, life expectancy (median life span from birth) was 89 days (~12%) greater in transgenic as compared with wild-type males and 112 days (~20%) greater in transgenic females (Fig. 4, A and C). Survival was assessed from a total of 57 females and 89 males; six Hcrt-UCP2 males remained alive at the time of this report and were treated as censored observations. Differential mortality was evaluated by Cox proportional-hazard regression with genotype, parent, and sex as main effects and sex-by-genotype as an interaction variable (tables S1 and S2). Each main effect was significant, but there was no effect of sex upon genotype, indicating that UCP2 overexpression in hypocretin neurons has an equal impact on the mortality of both genders.

Inspection of the complementary log-mortality plots between genotypes (Fig. 4, B and D) suggests that the ratio of their hazard rates is approximately constant with time. This assessment was verified by testing the significance of age as a time-dependent covariate; in males and females, there was no evidence to reject the hypothesis that the log mortality plots were parallel (table S3). The mortality rates for the two genotypes were proportional as required for Cox analysis. These data demonstrate that the reduction of CBT within a normal physiological range in Hcrt-UCP2 mice caused a parallel, proportional shift in the mortality rate trajectory, thus reducing the aging-related frailty of the mice (*22*). This demographic shift resembles the effects of CR upon mortality in mice and poikilotherms and differs from the effects of temperature reduction in poikilotherms, which changes the slope of mortality plots (*23*–*28*).

Although the mechanisms underlying the prolonged life span of Hcrt-UCP2 mice have yet to be elucidated, the aggregate data in several ways suggest that the mechanisms may be similar to those mediating the effects of CR. Although the reduction of CBT in Hcrt-UCP2 mice is small, metabolic requirements to maintain a lower CBT are reduced as demonstrated by the increased energy efficiency (Fig. 3C). This may lead to lower oxidative and free radical damage that, over the lifetime, ultimately prolonged life span.

We have shown that a modest and prolonged reduction of body core temperature can contribute to increased median life span in the absence of CR. The Hcrt-UCP2 "cool mice" represent a model for studying mechanisms underlying thermoregulation and metabolic regulation in mammals. They may also be useful for studying the effects of CBT on aging and longevity that are independent of the effects induced by CR.

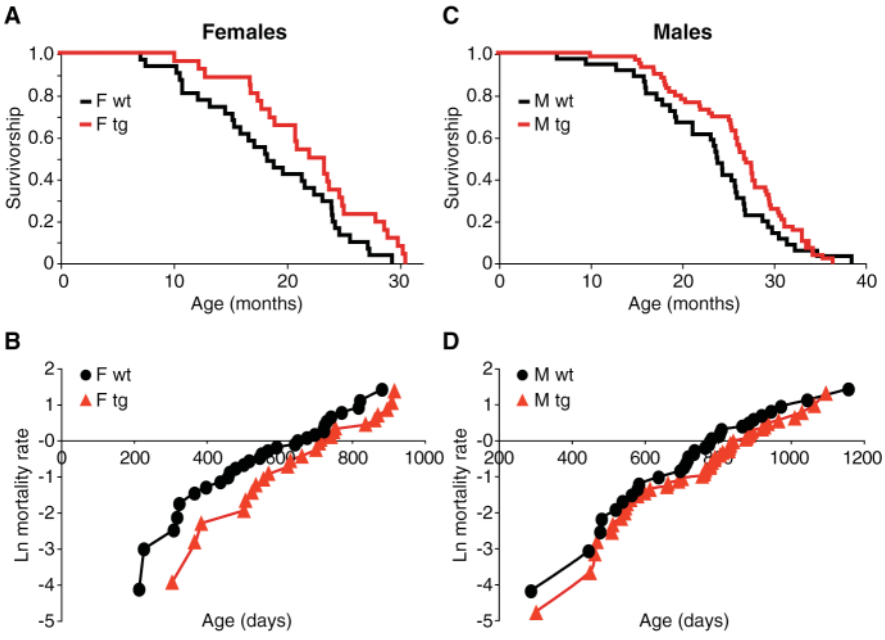


Fig. 4. Survival and mortality curves. Mice were fed ad libitum on sterilized breeder chow. Median life span was (**A**) 20% (females) and (**C**) 12% (males) greater in Hcrt-UCP2 mice relative to wild-type littermates. Complementary logarithmic plots (**B** and **D**) suggest that the ratio of the hazard rates for Hcrt-UCP2 and wild-type littermates is approximately constant with time. This assessment was verified by testing the significance of age as a time-dependent covariate (table S3). tg, transgenic; wt, wild-type.

References and Notes

1. J. A. Boulant, *Clin. Infect. Dis.* **31** (suppl. 5), S157 (2000).
2. D. Ricquier, F. Bouillaud, *Biochem. J.* **345**, 161 (2000).
3. J. G. Sutcliffe, L. de Lecea, *Nat. Rev. Neurosci.* **3**, 339 (2002).
4. J. T. Willie, R. M. Chemelli, C. M. Sinton, M. Yanagisawa, *Annu. Rev. Neurosci.* **24**, 429 (2001).
5. E. Mignot, S. Taheri, S. Nishino, *Nat. Neurosci.* **5** (suppl.), 1071 (2002).
6. L. de Lecea *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 322 (1998).
7. M. Jaszberenyi, E. Bujdoso, E. Kiss, I. Pataki, G. Telegdy, *Regul. Pept.* **104**, 55 (2002).
8. M. Monda, A. Viggiano, F. Fuccio, V. De Luca, *Brain Res.* **1018**, 265 (2004).
9. M. Szekely, E. Petervari, M. Balasko, I. Hernadi, B. Uzsoki, *Regul. Pept.* **104**, 47 (2002).
10. G. Yoshimichi, H. Yoshimatsu, T. Masaki, T. Sakata, *Exp. Biol. Med.* **226**, 468 (2001).
11. J. Hara *et al.*, *Neuron* **30**, 345 (2001).
12. H. C. Chen, Z. Ladha, S. J. Smith, R. V. Farese Jr., *Am. J. Physiol. Endocrinol. Metab.* **284**, E213 (2003).
13. R. Liu, R. Walford, *Nature* **212**, 1277 (1966).
14. C. McCay, M. Cromwell, L. Maynard, *J. Nutr.* **10**, 63 (1935).
15. R. Weindruch, R. L. Walford, S. Fligiel, D. Guthrie, *J. Nutr.* **116**, 641 (1986).
16. P. H. Duffy, J. E. Leakey, J. L. Pipkin, A. Turturro, R. W. Hart, *Environ. Res.* **73**, 242 (1997).
17. A. Koizumi, M. Tsukada, Y. Wada, H. Masuda, R. Weindruch, *J. Nutr.* **122**, 1446 (1992).
18. W. Duan, M. P. Mattson, *J. Neurosci. Res.* **57**, 195 (1999).
19. D. K. Ingram, R. Weindruch, E. L. Spangler, J. R. Freeman, R. L. Walford, *J. Gerontol.* **42**, 78 (1987).
20. M. P. Mattson, *Brain Res.* **886**, 47 (2000).
21. H. Zhu, Q. Guo, M. P. Mattson, *Brain Res.* **842**, 224 (1999).
22. J. W. Vaupel, K. G. Manton, E. Stallard, *Demography* **16**, 439 (1979).
23. L. Partridge, S. D. Pletcher, W. Mair, *Mech. Ageing Dev.* **126**, 35 (2005).
24. W. Mair, P. Goymer, S. D. Pletcher, L. Partridge, *Science* **301**, 1731 (2003).
25. J. Miquel, P. R. Lundgren, K. G. Bensch, H. Atlan, *Mech. Ageing Dev.* **5**, 347 (1976).
26. M. Parmar, D. Machin, Eds., *Survival Analysis: A Practical Approach* (Wiley, New York, 1995).
27. S. D. Pletcher *et al.*, *Curr. Biol.* **12**, 712 (2002).
28. M. Tatar, *Senescence*, in C. W. Fox, D. A. Roff, D. J. Fairbain, Eds., *Evolutionary Ecology: Concepts and Case Studies* (Oxford Univ. Press, Oxford, 2001), pp. 128–141.

29. S. N. Austad, D. M. Kristan, *Aging Cell* **2**, 201 (2003).
30. We thank J. Buxbaum, M. Behrens, M. Brennan, E. Crawford, C. Davis, K. Eusko, R. Fernandez, J. Koziol, M. Martes, T. Hurtado de Mendoza, J. Pemberton, M. Sabbatini, V. Zhukov, and D. Wills for their input and assistance; T. Sakurai for providing the orexin ataxin 3 mice; and M. Tatar for assistance with survival analysis. This work was supported by the Harold Dorris Neurological Research Institute, The Ellison Medical Foundation, NS043501 and MH58543.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/825/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 to S3
References

7 July 2006; accepted 8 September 2006
10.1126/science.1132191

Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex

Daria Knoch,[1,2,3]* Alvaro Pascual-Leone,[4] Kaspar Meyer,[1] Valerie Treyer,[5] Ernst Fehr[1,3]*

Humans restrain self-interest with moral and social values. They are the only species known to exhibit reciprocal fairness, which implies the punishment of other individuals' unfair behaviors, even if it hurts the punisher's economic self-interest. Reciprocal fairness has been demonstrated in the Ultimatum Game, where players often reject their bargaining partner's unfair offers. Despite progress in recent years, however, little is known about how the human brain limits the impact of selfish motives and implements fair behavior. Here we show that disruption of the right, but not the left, dorsolateral prefrontal cortex (DLPFC) by low-frequency repetitive transcranial magnetic stimulation substantially reduces subjects' willingness to reject their partners' intentionally unfair offers, which suggests that subjects are less able to resist the economic temptation to accept these offers. Importantly, however, subjects still judge such offers as very unfair, which indicates that the right DLPFC plays a key role in the implementation of fairness-related behaviors.

Across species, humans have been spectacularly successful in limiting the impact of self-interest even in interactions between genetically unrelated strangers, by developing and enforcing social norms (*1*, *2*). Fairness norms, in particular, play a crucial role in social life across many cultures (*3*). They are enforced by reciprocally fair behaviors, which imply that kind acts are reciprocated with kindness, whereas hostile or unfair acts are reciprocated with hostility (*4*). The role of fairness also has been acknowledged in formal theories of reciprocal fairness (henceforth, "reciprocity") (*4*) and inequity aversion (*5*), both of which assume that people trade off fairness goals against the goal of increasing their material resources.

The Ultimatum Game (*6*) illustrates the tension between economic self-interest, on the one hand, and reciprocity and equity motives, on the other. In this game, two anonymous individuals, a proposer and a responder, have to agree on the division of a given amount of money, say $20, according to the following rules: The proposer can make exactly one suggestion on how the $20 should be allocated between the two by making an integer offer X to the responder. Then the responders can either accept or reject X. In case of a rejection, both players earn $0; in case of acceptance, the responder earns X and the proposer earns $20 – X. If economic self-interest alone motivates the responder, he will accept even a very low offer, say $1, because $1 is better than $0. However, if concerns for reciprocity and equity motivate him, he might reject low offers because he views them as insultingly unfair and inequitable. The responder thus faces a conflict in case of low offers between his economic self-interest, which encourages him to accept the offer, and his fairness goals, which drive him toward rejecting it.

Strong evidence (*3*, *7*) suggests that many people reject low offers in the game, even if stake levels are as high as 3 months' income (*8*). Rejection rates up to 80% have been observed (*7*) for offers below 25% of the available money, and a pioneering imaging study (*9*) showed that both the anterior insula—an important brain area involved in the processing of emotions (*10*)—and the dorsolateral prefrontal cortex (DLPFC) are activated when responders decide whether to accept or reject an unfair offer. The fact that both the right and left DLPFCs are more strongly activated when subjects face unfair offers compared with when they face fair offers is of particular interest for our purposes. The DLPFC is widely thought to be involved in executive control, goal maintenance, and the inhibition of prepotent responses (*11*). All of these functions are relevant for the responder in the Ultimatum Game, because there are likely to be several competing goals—fairness goals and self-interest—and the questions are as follows: Which of them should be maintained, i.e., given priority, and which motivational impulse should be restrained?

One plausible hypothesis about the role of the DLPFC is that unfair offers generate an impulse to reject, and that DLPFC activity is involved in controlling this impulse (*9*). According to this hypothesis, DLPFC activity is involved in the cognitive control of the emotional impulse associated with fairness goals. A contrasting but equally plausible hypothesis is that fundamental impulses associated with self-interest need to be controlled in order to maintain and to implement culture-dependent fairness goals (*3*). According to this hypothesis, the DLPFC is involved in overriding selfish impulses, the latter of which may also be strongly associated with emotional forces. This second hypothesis has a "Freudian" flavor because Freud's theory of the ego and the superego is based on the assumption that the ego and the superego need to override the fundamentally selfish nature of the id in order for human beings to behave in reasonable and morally appropriate ways (*12*). Both hypothesis are, however, consistent with dual-systems approaches (*13*–*16*) that stress the fundamental differences between emotional (impulsive) and deliberative systems, or between automatic and controlled processes, because both fairness and selfish impulses may have strong emotional content.

If we suggest that DLPFC may be involved in inhibiting or overriding self-interest motives, we do not want to imply that DLPFC directly suppresses other brain areas that represent self-interest. Rather the term "inhibition" is a convenient short term for top-down control (or executive control), whose overall effect is a reduction in the weight

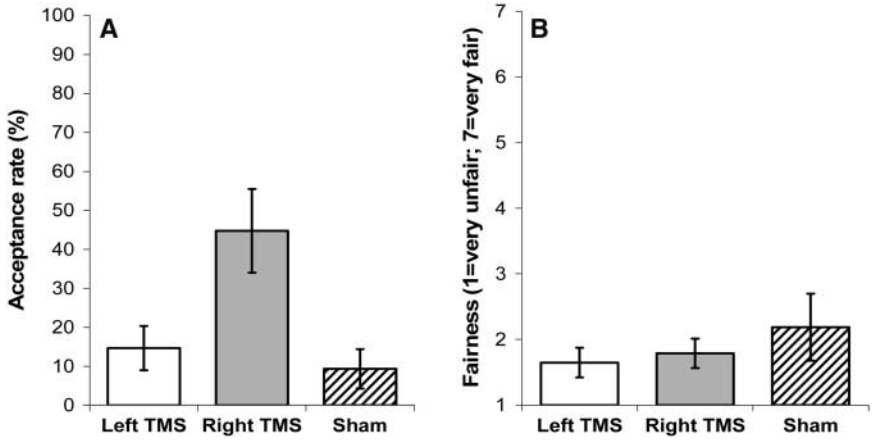


Fig. 1. Behavioral responses and fairness judgments (means ± SEM) related to the most unfair offer of CHF 4 in the human offer condition. (**A**) Acceptance rates across treatment groups. Subjects whose right DLPFC is disrupted exhibit a much higher acceptance rate than those in the other two treatment groups (Mann-Whitney U tests, two-tailed, $P < 0.05$). (**B**) Perceived unfairness across treatments (1 = very unfair; 7 = very fair). Subjects in all three treatment groups perceive an offer of 4 as very unfair, and there are no significant differences across groups.

[1]Institute for Empirical Research in Economics, University of Zurich, Blümlisalpstrasse 10, 8006 Zurich, Switzerland. [2]Department of Neurology, University Hospital Zurich, Switzerland. [3]Collegium Helveticum, Schmelzbergstrasse 25, 8092 Zurich, Switzerland. [4]Center for Noninvasive Brain Stimulation, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, MA 02115, USA. [5]PET Center, Nuclear Medicine, Department of Radiology, University Hospital, 8091 Zurich, Switzerland.

*To whom correspondence should be addressed. E-mail: dknoch@iew.unizh.ch; efehr@iew.unizh.ch

of self-interested impulses on an individual's action. Thus, rather than directly suppressing neural activities that represent self-interested impulses, the DLPFC may be part of a network that modulates the relative impact of fairness motives and self-interest goals on decision-making. The final outcome of this modulation may then be a weakening of the impact of self-interest motives on decision-making.

The fact that the DLPFC is more strongly activated with unfair offers than with fair ones (*9*) does not necessarily mean that this brain region is crucially involved in the implementation of fair behaviors. In principle, it is even possible that DLPFC activation is not causally involved in the decision to accept or reject unfair offers, but instead, represents only a by-product of some other process. In order to examine whether DLPFC activity is crucial in the responders' decisions and to discriminate between the two hypotheses mentioned above, we applied low-frequency repetitive transcranial magnetic stimulation (rTMS) to 52 subjects in the role of the responder in an anonymous Ultimatum Game with a stake size of Swiss francs (CHF) 20 (CHF 1 ≈ $0.80). We limited the proposer's strategy space by only permitting offers of CHF 10, 8, 6, or 4 (*17*) in order to generate enough observations on the responders' side. CHF 10 is obviously the fairest offer, because it splits the stake size equally, whereas CHF 4 is the most unfair offer. Each responder played the Ultimatum Game 20 times with 20 different anonymous partners. In order to investigate a possible hemispheric laterality in the role of DLPFC on responders' decisions, we applied low-frequency rTMS for 15 min to the right (19 subjects) or to the left DLPFC (17 subjects). Low-frequency rTMS for the duration of several minutes leads to a suppression of activity in the stimulated brain region that outlasts the duration of the rTMS train for about half the duration of the stimulation. The existence of a group that receives rTMS to the right DLPFC and a control group that receives rTMS to the left DLPFC is also important because this controls for the potential side effects of rTMS (*18*, *19*), including discomfort, irritation, and mood changes. In addition, we had a further control condition where we applied sham stimulation for 15 min to the right or left DLPFC (16 subjects). Each subject participated in only one of the three conditions (left stimulation, right stimulation, or sham), and none had experienced TMS previously. This is important because subjects who experienced real rTMS before or after sham stimulation on the same day are very likely to be able to distinguish between the two stimulations (*17*), which questions the control status of the sham stimulation. Therefore, a pioneering attempt (*20*), which lacked an active rTMS control stimulation and an across-subject sham control, could not attribute possible behavioral changes associated with low-frequency rTMS to the disruption of DLPFC.

How will the disruption of DLPFC with low-frequency rTMS affect the responders' behavior? If the DLPFC is involved in implementing fair behavior, which requires overriding selfish impulses, disrupting this brain region should increase the acceptance rate for unfair offers relative to the sham-stimulation condition. In other words, if we disrupt activity in a brain region hypothesized to place controls on selfish impulses, we should functionally weaken the control, and selfish impulses should thus have a stronger impact on decision-making; the acceptance rate of unfair offers should, therefore, increase. Alternatively, if DLPFC activity is involved in the cognitive control associated with the inhibition of fairness impulses, low-frequency rTMS applied to this brain area should reduce the acceptance rate of unfair offers, because the fairness impulses should affect behavior more strongly if the ability to inhibit them is reduced. Thus, the two hypotheses make opposite predictions on how low-frequency rTMS of DLPFC affects acceptance rates relative to the sham stimulation.

The application of rTMS to the right and left DLPFC also enables us to test a lateralization hypothesis. A number of studies have reported preferential right-hemispheric involvement in the inhibitory control of behavior (*21*–*24*). This has been shown in go/no-go tasks (*21*, *22*) and risky choice tasks (*24*), for example, and there appears to be a right hemispheric lateralization of syndromes, in which impairments in decision-making and social behavior seem to reflect a breakdown of control processes (*25*). We, therefore, conjectured that disruption of the right DLPFC might be associated with different effects compared with the disruption of the left DLPFC.

In addition to the conditions (left DLPFC, right DLPFC, or sham) mentioned above, we also implemented the following treatment variation. In each of the three conditions, a responder played 10 games with partners who were responsible for their offers, because they decided how much to offer (human-offer condition), as well as 10 games with partners who where not responsible for their offers because a computer randomly generated the offers (computer-offer condition). The latter condition is interesting because previous evidence indicates (*26*) that two fairness motives—reciprocity (*4*) and inequity aversion (*5*)—are simultaneously activated in the human-offer condition, whereas only one fairness motive—inequity aversion—is operative in the computer-offer condition. A responder motivated by reciprocity rejects a low offer because he wants to punish the proposer's unfair behavior, whereas the inequity aversion motive is characterized by resistance against the unfair distribution of income that results from a low offer, no matter how this distribution was generated. Therefore, the motive to punish the partner for an unfair offer cannot play a role in the computer-offer condition, because the partner is not responsible for it. As a consequence, the behavioral impact of rTMS in the computer-offer condition enables us to specify the interpretation of our results by comparing the effects of the disruption of the DLPFC in situations where the reciprocity motive is present and absent.

As expected, the acceptance rates varied strongly across offers. In the human-offer condition, offers of 4 were accepted on average in 24% of the trials, whereas the acceptance rate for offers of 6 was 67%, and offers of 8 were accepted in 99% of the cases. For our purposes, acceptance behavior with regard to the lowest offers is most interesting, because the tension between fairness and self-interest is greatest in this case. After sham rTMS, the acceptance rate for the most unfair offer was 9.3% and after real rTMS of the left DLPC it was 14.7% (Fig. 1A). These results contrast sharply with the acceptance rate of 44.7% after rTMS of the right DLPFC. In fact, 37% of the subjects accepted all unfair offers after right DLPFC disruption, whereas no subjects in the sham or the left DLPFC group accepted all of them. The differences across all three groups are significant (Kruskal-Wallis test, $H = 7.265$, df = 2, $n = 52$, $P = 0.026$), and pairwise two-tailed Mann-Whitney U tests confirm that the right DLPFC group has a significantly higher acceptance rate

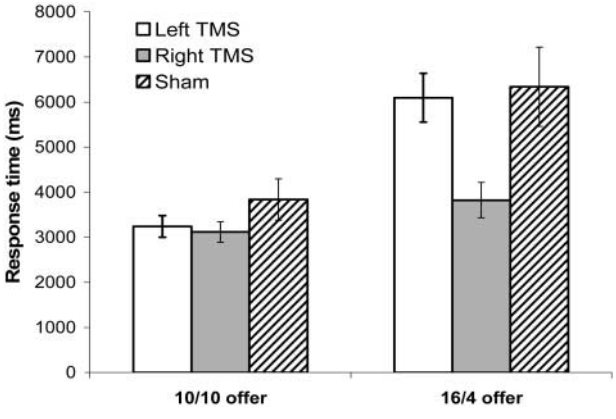


Fig. 2. Response times for fair and unfair offers in the human offer condition. If subjects face a fair offer of 10 they quickly accept the offer, and there are no significant differences across treatment groups (Kruskal-Wallis test, $P = 0.264$). If subjects face an unfair offer of 4 the response time strongly increases for subjects whose left DLPFC is disrupted and for those who receive sham stimulation. However, subjects whose right DLPFC is disrupted accept unfair offers almost as quickly as fair offers, leading to strong response-time differences across treatments for the case of unfair offers (Kruskal-Wallis test, $P = 0.028$)

than either the left DLPFC ($Z = -1.969$, $n = 36$, $P = 0.049$) or the sham group ($Z = -2.388$, $n = 35$, $P = 0.017$). The same results hold if we pool the behavioral responses to unfair offers of 4 and 6 and conduct a repeated measures analysis of variance (ANOVA) of treatment (left DLPFC, right DLPFC, or sham) × offer (4, 6). We find a main effect of treatment [$F(42,49) = 5.30$, $P = 0.008$], and Fisher's post hoc test demonstrates that subjects who received rTMS over right DLPFC were more likely to accept offers of 4 or 6 than those stimulated over left DLPFC ($P = 0.041$) or those who received sham rTMS ($P = 0.003$). Interestingly, we found no interaction effect of treatment × offer ($P = 0.398$), which indicated that subjects' acceptance behavior shows a similar pattern for both unfair offers (i.e., higher acceptance rate after right DLPFC disruption compared with the other two conditions).

These differences across conditions can neither be attributed to different propensities to behave reciprocally nor to differences in individual impulsivity across treatment groups. Roughly 10 days after the experiment, the subjects also completed personality questionnaires that assessed their impulsivity (*27*) and propensity to reciprocate (*17*, *28*). We observed no differences across treatment groups for either impulsivity (Kruskal-Wallis test, $H = 3.693$, df = 2, $P = 0.158$) or reciprocity (Kruskal-Wallis test, $H = 0.853$, df = 2, $P = 0.653$). Moreover, the treatment differences in acceptance rates remain highly significant if we control for individuals' impulsivity and reciprocity scores in a repeated measures ANOVA [main effect of treatment ($F = 4.701$, $P = 0.014$); Fisher's post hoc test: right versus left DLPFC ($P = 0.020$), right versus sham ($P = 0.007$)].

Interestingly, although rTMS of the right DLPFC reduced the rejection rate, rTMS did not change subjects' fairness judgments. We elicited subjects' fairness judgments with regard to different offers on a seven-point scale (*17*) immediately after the Ultimatum Game experiment. Subjects in all three treatment groups judged the lowest offer of 4 as rather unfair when the human partner made this offer (Fig. 1B), and a Kruskal-Wallis test indicated no differences in fairness judgments across treatments ($H = 0.197$, df = 2, $P = 0.911$). Similarly, there were also no differences in the fairness assessment of offers of 6 across treatments ($H = 0.487$, df = 2, $P = 0.784$). Thus, disruption of the right DLPFC diminishes fair behavioral responses to low offers but does not affect subjects' fairness judgments.

The results reported above support the hypothesis that right, but not left, DLPFC activity, is crucial for the ability to override selfish impulses in order to reject offers perceived as unfair. The response-time difference for accepted unfair and fair offers across groups provides further support for this hypothesis (Fig. 2). If subjects face a fair offer of 10 they quickly accept, and no response-time differences across treatment groups are observed (Kruskal-Wallis test, $H = 2.662$, df = 2, $P = 0.264$). In sharp contrast, subjects who receive rTMS to the left DLPFC or sham stimulation need much longer to accept unfair offers of 4 than subjects who receive fair offers, which suggests that there is a conflict between self-interest and fairness motives at unfair offers that causes an increase in response time. However, subjects whose right DLPFC is disrupted exhibit similar response times for both fair and unfair offers (Fig. 2). Thus, large and significant differences across treatment groups occurred at offers of 4 (Kruskal-Wallis test, $H = 8.051$, df = 2, $P = 0.0179$); subjects in the right DLPFC group accepted unfair offers significantly faster than do subjects in the left DLPFC group (Mann-Whitney U test, two-tailed, $P = 0.018$) or in the sham group (Mann-Whitney U test, two-tailed, $P = 0.028$). In terms of response time, subjects with right DLPFC disruption seem to be less able to resist the selfish temptation to accept low offers although they view them as unfair.

The hypothesis that right DLPFC is crucial in implementing fairness behaviors by overriding self-interested impulses also has implications for the computer-offer condition. Recall that the motive for punishing the partner for unfair offers cannot play a role in this condition, because the partner is not responsible for the offers. As a consequence, the fairness forces are weak, and therefore, the effects of disrupting the ability to implement fair behaviors should also be so. This pattern contrasts with the human-offer condition where the fairness forces are strong; disruption of the ability to implement fair actions should, therefore, have a strong impact on behavior. Thus, the hypothesis that right DLPFC is crucial for the ability to implement fairness goals predicts that disruption of the right DLPFC activity will generate weaker effects in the computer-offer condition than in the human-offer condition.

The data indeed support this prediction. The average acceptance rate of an offer of 4 was 67.3% in the computer-offer condition, which was much higher than the 24% in the human-offer condition. The fact that fairness forces were weaker in the computer-offer condition is also indicated by the perceived unfairness of an offer of 4 (compare Fig. 3B with Fig. 1B) but, as in the human-offer condition, there were no differences in the perceived unfairness across treatment groups (Kruskal-Wallis test, $H = 0.052$, $P = 0.974$). In contrast to the human-offer condition there were no significant differences in acceptance rates across left DLPFC, right DLPFC, and sham condition (Fig. 3A; Kruskal-Wallis test, $H = 2.370$, $P = 0.306$), which suggests that the disruption of right DLPFC has no, or only minor, effects in the computer-offer condition. A similar conclusion was suggested by a repeated measures ANOVA, which indicated an interaction between treatment type (i.e., human- versus computer-offer condition) and treatment group (i.e., left DLPFC, right DLPFC, or sham condition) ($F = 3.318$, $P = 0.045$). This shows that the computer-offer condition significantly reduces the behavioral differences across treatment groups relative to the human-offer condition.

These findings illustrate the importance of rTMS for progress in understanding the neural basis of human decision-making. First, previous neuroimaging studies (*9*) put forward the view that the DLPFC is crucial in overriding fairness impulses when self-interest and fairness motives are in conflict. However, our data indicate that the DLPFC plays a key role in overriding or weakening self-interested impulses and thus enables subjects to implement their fairness goals. Subjects are much more willing to behave selfishly, i.e., to accept unfair offers, after disruption of the right DLPFC. In addition, they accept unfair offers almost as quickly as fair of-

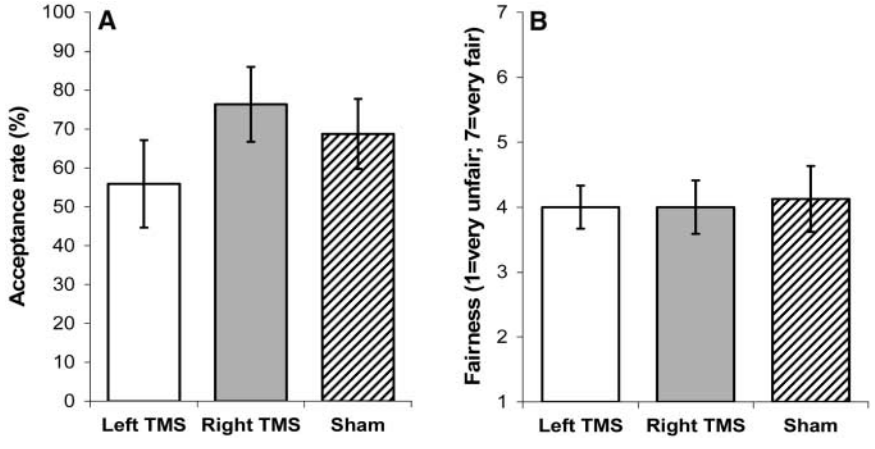


Fig. 3. Behavioral responses and fairness judgments (means ± SEM) related to the most unfair offer of CHF 4 in the computer offer condition. (**A**) Acceptance rates across treatment groups. The differences across treatment groups are not significant, and they are significantly smaller than in the human offer condition (repeated measures ANOVA, $P < 0.05$). (**B**) Perceived unfairness across treatments (1 = very unfair; 7 = very fair). Subjects in all three treatment groups perceive an offer of 4 as less unfair than in the human offer condition (compare with Fig. 1B), and there are no differences in fairness judgments across treatment groups.

fers, which suggests that self-interest impulses have a stronger impact on behavior. Second, our rTMS study also enables us to claim a causal role of DLPFC activity in the implementation of fairness motives when self-interest and fairness are in conflict. The only previous attempt (*20*) could not attribute a weak behavioral effect of rTMS to the disruption of DLPFC because it used a confounded sham control and lacked an active rTMS control stimulation (*29*). Third, disruption of right DLPFC activity only affects fairness-related behaviors but not fairness judgments. Subjects behave as if they can no longer implement their fairness goals after disruption of the right DLPFC. This finding is also interesting in light of evidence suggesting that patients with right prefrontal lesions are characterized by the inability to behave in normatively appropriate ways, despite the fact that they have the judgment necessary for normative behavior (*30*); the findings thus support the importance of right prefrontal areas for normatively appropriate behaviors. Thus, a dysfunction of the right DLPFC, or its specific connections, may underlay certain psychopathological disorders that are characterized by excessive selfish tendencies and a failure to obey basic social norms. Fourth, the fact that there is no behavioral effect of right DLPFC disruption in the computer offer condition, where the reciprocity motive is absent, supports the role of the right DLPFC in the implementation of reciprocally fair behaviors. Fifth, the data show that only the right, but not the left, DLPFC activity plays a causal role, because disruption of the left DLPFC causes no behavioral changes relative to the sham control, whereas disruption of the right DLPFC sharply increases the acceptance rate of unfair offers. This fact is particularly interesting in the light of neuroimaging data (*9*) that show that left DLPFC is activated if subjects face unfair offers, raising exciting questions about its exact role and the possible interplay of the left and the right hemispheres in the implementation of fairness-related behaviors that require an overriding of selfish impulses. Finally, the reported findings provide evidence for theoretical approaches (*13*–*16*) to social cognition and decision-making that stress the fundamental role of DLPFC in neural networks that support deliberative processes in human decision-making.

References and Notes

1. R. T. Boyd, P. Richerson, *The Origin and Evolution of Cultures* (Oxford Univ. Press, Oxford, 2005).
2. E. Fehr, U. Fischbacher, *Nature* **425**, 785 (2003).
3. J. Henrich *et al.*, *Am. Econ. Rev.* **91**, 73 (2001).
4. M. Rabin, *Am. Econ. Rev.* **83**, 1281 (1993).
5. E. Fehr, K. M. Schmidt, *Q. J. Econ.* **114**, 817 (1999).
6. W. Güth, R. Schmittberger, B. Schwarze, *J. Econ. Behav. Organ.* **3**, 367 (1982).
7. C. F. Camerer, *Behavioral Game Theory—Experiments in Strategic Interaction* (Princeton Univ. Press, Princeton, NJ, 2003).
8. L. A. Cameron, *Econ. Inq.* **37**, 47 (1999).
9. A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, J. D. Cohen, *Science* **300**, 1755 (2003).
10. A. J. Calder, A. D. Lawrence, A. W. Young, *Nat. Rev. Neurosci.* **2**, 352 (2001).
11. E. K. Miller, J. D. Cohen, *Annu. Rev. Neurosci.* **24**, 167 (2001).
12. S. Freud, *The Ego and the Id* (Norton, New York, 1962).
13. A. Bechara, *Nat. Neurosci.* **8**, 1458 (2005).
14. A. B. Satpute, M. D. Lieberman, *Brain Res.* **1079**, 86 (2006).
15. A. G. Sanfey, G. Loewenstein, S. M. McClure, J. D. Cohen, *Trends Cogn. Sci.* **10**, 108 (2006).
16. G. F. Loewenstein, T. O'Donoghue, "Animal spirits: Affective and deliberative processes in economic behavior" (Social Science Research Network, 2004), available at SSRN http://ssrn.com/abstract=539843.
17. Materials and methods are available as supporting material on *Science* Online.
18. E. M. Robertson, H. Theoret, A. Pascual-Leone, *J. Cogn. Neurosci.* **15**, 948 (2003).
19. B. Abler *et al.*, *Brain Topogr.* **17**, 193 (2005).
20. M. van 't Wout, R. S. Kahn, A. G. Sanfey, A. Aleman, *Neuroreport* **16**, 1849 (2005).
21. A. R. Aron, P. C. Fletcher, E. T. Bullmore, B. J. Sahakian, T. W. Robbins, *Nat. Neurosci.* **6**, 115 (2003).
22. H. Garavan, T. J. Ross, E. A. Stein, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8301 (1999).
23. A. R. Aron, T. W. Robbins, R. A. Poldrack, *Trends Cogn. Sci.* **8**, 170 (2004).
24. D. Knoch *et al.*, *J. Neurosci.* **26**, 6469 (2006).
25. S. E. Starkstein, R. G. Robinson, *J. Nerv. Ment. Dis.* **185**, 108 (1997).
26. S. Blount, *Organ. Behav. Hum. Decis. Process.* **63**, 131 (1995).
27. E. S. Barratt, M. S. Stanford, in *Personality Characteristics of the Personality Disorders*, C. G. Costello, Ed. (Wiley, New York, 1996), pp. 91–119.
28. M. Perugini, M. Callucci, F. Presaghi, A. P. Ercolani, *Eur. J. Personality* **17**, 251 (2003).
29. This study does not control for repeated measurements and finds no significant behavioral effect at the 5% level. This may be due to the fact that only seven subjects participated and that the two treatments (sham and the right DLPFC condition) were only 30 min apart.
30. A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain* (Harper Collins, New York, 1995).

31. This paper is part of the research priority program at the University of Zurich on the "Foundations of Human Social Behavior—Altruism versus Egoism" and the research program of the Collegium Helveticum on the emotional foundations of moral behavior which is supported by the Cogito Foundation. D.K. also acknowledges support from the Swiss National Science Foundation and the UBS AG (donation). A.P.-L. is supported in part by grants from the National Institutes of Health (K24 RR018875) and the Center for Integration of Medicine and Innovative Technology (CIMIT).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1129156/DC1
Materials and Methods
Tables S1 to S3
References

25 April 2006; accepted 22 September 2006
Published online 5 October 2006;
10.1126/science.1129156
Include this information when citing this paper.

The Polarity Protein Par-3 Directly Interacts with p75^{NTR} to Regulate Myelination

Jonah R. Chan,[1]* Christine Jolicoeur,[2] Junji Yamauchi,[3] Jimmy Elliott,[2] James P. Fawcett,[4]‡ Benjamin K. Ng,[1] Michel Cayouette[2,5]*

Cell polarity is critical in various cellular processes ranging from cell migration to asymmetric cell division and axon and dendrite specification. Similarly, myelination by Schwann cells is polarized, but the mechanisms involved remain unclear. Here, we show that the polarity protein Par-3 localizes asymmetrically in Schwann cells at the axon-glial junction and that disruption of Par-3 localization, by overexpression and knockdown, inhibits myelination. Additionally, we show that Par-3 directly associates and recruits the p75 neurotrophin receptor to the axon-glial junction, forming a complex necessary for myelination. Together, these results point to a critical role in the establishment of cell polarity for myelination.

The myelin sheath is a specialized membrane component in the vertebrate nervous system that is essential for the optimal transmission of neuronal action potentials. In the peripheral nervous system, Schwann cells (SC) are responsible for myelinating axons. Recently, environmental signals, particularly the neuregulins (*1*) and the neurotrophins (*2*, *3*), have been shown to regulate SC myelination. Specifically, neurotrophin 3 (NT-3) promotes SC migration and inhibits myelination (*2*–*4*). In contrast, brain-derived neurotrophic factor (BDNF) inhibits SC migration and promotes myelination through the p75 neurotrophin receptor (NTR) (*2*, *3*, *5*). Much less is known, however, about the intrinsic mechanisms governing SC myelination. The formation of myelin by SCs is a highly polarized process, which consists of the unidirectional wrapping of multiple layers of membrane concentrically around an axon, initiated exclusively at the site of the axon-glial junction, raising the question of what regulates the asymmetric initiation of myelination.

To address this question, we examined the distribution of Par-3, a member of the Par family of adaptor proteins involved in the establishment of cell polarity in various cellular contexts (*6*–*15*) in SC/dorsal root ganglion (DRG) neuronal cocultures (Fig. 1A) and in sciatic nerves (Fig. 1B). Interestingly, we found that Par-3 is enriched asymmetrically at the membrane of

premyelinating SCs (Fig. 1, A and B) and did not colocalize with markers for the Golgi or the endoplasmic reticulum (fig. S1). This asymmetric localization of Par-3 formed a bandlike pattern at the site of contact between the Schwann cell and the axon, as revealed most clearly when staining for neurofilament to reveal the axons, which did not express detectable amounts of Par-3 (Fig. 1, C and D). To better visualize the axon-glial junction, we stained for N-cadherin, an adhesion protein enriched at the SC/axon junction (*16*). In the SC/DRG cocultures, we found that Par-3 and N-cadherin colocalize (Fig. 1E). Similar to previous findings with N-cadherin, the asymmetric localization of Par-3 was also identified in SCs cultured without neurons, but only when SCs made contact with one another (Fig. 1F, arrow). In the absence of contact, Par-3 localization was diffuse and uniform (Fig. 1F, arrowhead). Thus, it appears that the redistribution of Par-3 is a direct result of cell-cell contact, which can be either SC-SC or SC-axon contact, consistent with previous studies showing that cell-cell contact is necessary for asymmetric localization of Par proteins (*10*, *11*, *15*, *17*, *18*). To visualize the ultrastructural localization of Par-3 at the axon-glial junction, immunogold electron microscopy (EM) was performed on premyelinating SC/DRG cocultures. The asymmetric localization of Par-3 was identified by the electron-dense gold particles deposited in the SC at the site of the SC/axon junction in cross section (Fig. 1G) and in longitudinal section (Fig. 1H).

Based on these results, we hypothesized that the asymmetric distribution of Par-3 at the axon-glial junction might be required to initiate myelin formation by SCs. If this were the case, disruption of Par-3 localization by either overexpression or knockdown in SCs should inhibit myelination. We used a well-established SC/DRG coculture system and a retroviral expression system to manipulate Par-3 expression and assess the ability of SCs to myelinate DRG axons (see supporting online material). Because retroviral vectors infect only actively dividing cells, this method ensures that SCs are infected whereas the DRG neurons remain uninfected. When we overexpressed Par-3 in SCs, we disrupted the polarized asymmetric localization of Par-3 (fig. S2), and SCs were unable to myelinate DRG axons, as revealed by staining for myelin basic protein (MBP) (Fig. 2A) and other myelin markers. Noninfected SCs on the same axon formed myelin normally (Fig. 2A). We performed multiple experiments and could not identify any Par-3–overexpressing SCs forming mature myelin. Interestingly, Par-3–overexpressing SCs aligned on axons normally, as observed by phase contrast microscopy, and formed what looked like internodes (Fig. 2A, arrows).

We next asked whether Par-3 was required for myelin formation by SCs. We generated retroviral vectors expressing both green fluorescent protein (GFP) and a short-hairpin RNA (shRNA) sequence for Par-3. When tested in SCs, the Par-3 shRNA significantly reduced Par-3 expression (Fig. 2, B and C), whereas a control shRNA directed against another region of Par-3 did not alter expression and neither did a control retroviral vector encoding GFP only (Fig. 2B). We infected SCs with these retroviral vectors and tested the ability of SCs to myelinate DRG axons. Although SCs infected with the different control shRNA formed myelin normally, knocking down Par-3 expression in SCs significantly reduced their ability to myelinate (Fig. 2, D and E) without affecting proliferation (fig. S3). In some cases, SCs appeared to synthesize myelin proteins as seen in Fig. 2D (open arrow) but could not adequately myelinate the axons. Similar results were obtained with three additional nonoverlapping shRNA constructs previously reported to

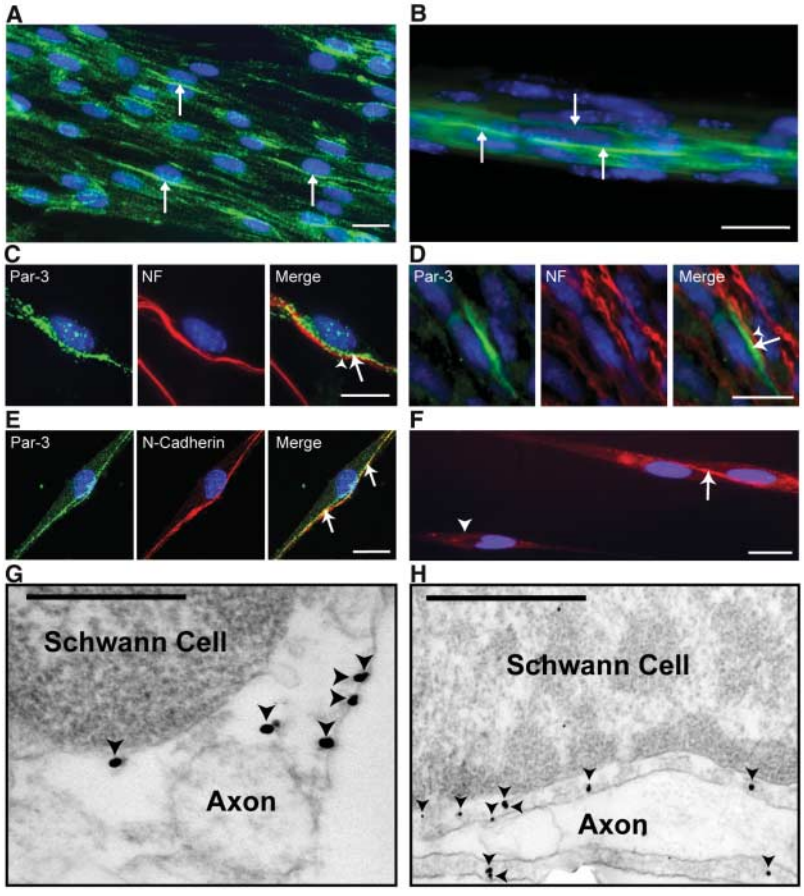


Fig. 1. Asymmetric localization of Par-3 in SCs at the axon/glial junction. (**A** and **B**) Immunostaining for Par-3 protein (green) in premyelinating SCs cultured with DRG neurons (A) and in teased fibers of embryonic rat sciatic nerve at 18 days gestation (B). Note the asymmetric localization of Par-3 (arrows). (**C** and **D**) Par-3 protein (green, arrows) and neurofilament (red, arrowheads) in SC/DRG neuron cocultures (C) and in sciatic nerve sections from embryonic rats at 18 days gestation (D). Par-3 and neurofilament are juxtaposed to one another and Par-3 is absent in the axons. (**E**) Immunostaining for Par-3 (green) and N-cadherin (red) in SC/DRG neuron cocultures. The arrows point to the colocalization (yellow). (**F**) SCs cultured in the absence of neurons display asymmetric localization of Par-3 (red, arrow) only upon contact with other SCs. In the absence of SC-SC contact, diffuse and uniform Par-3 staining is observed (arrowhead). Nuclei were stained using Hoechst (blue). Scale bars, 10 μm. (**G**) Immunogold EM on premyelinating SC/DRG cocultures in cross section. Par-3 is enriched in SCs at the site of axon contact, as identified by the electron-dense gold particles (arrowheads). (**H**) A longitudinal section of a premyelinating SC and axon. Par-3 localization is enriched in the SC membrane ensheathing the axon, as identified by the electron-dense gold particles (arrowheads). Scale bars, 500 nm.

[1]Department of Cell and Neurobiology, Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles 90089, USA. [2]Cellular Neurobiology Laboratory, Institut de Recherches Cliniques de Montreal (IRCM), Montreal, QC, Canada. [3]Department of Molecular Pharmacology, National Research Institute of Child Health and Development, Tokyo, Japan. [4]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada. [5]Département de Médecine, Université de Montréal.

*To whom correspondence should be addressed. E-mail: michel.cayouette@ircm.qc.ca (M.C.); jonah.chan@usc.edu (J.R.C.)

‡Present address: Departments of Pharmacology and Surgery, Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada.

knock down Par-3 expression (*19*) (fig. S4). Importantly, disruption of Par-3 expression did not affect axon/glial junction formation, as revealed by N-cadherin staining (fig. S5). These results indicate that disrupting the asymmetric localization of Par-3 in premyelinating SCs by either overexpression or knockdown inhibits myelination, suggesting that asymmetric localization of Par-3 may be necessary for myelination.

How does a polarity protein such as Par-3 regulate myelination? Previous studies have shown that BDNF binds $p75^{NTR}$ to inhibit SC migration and promote myelination (*3*, *5*). Consistent with these results, $p75^{NTR}$ knockout mice display a delay in myelination, thinner myelin sheaths, and an overall reduction in myelinated fibers in adulthood (*3*). Because Par-3 is a well-known scaffolding protein, it raises the possibility that Par-3 and $p75^{NTR}$ might form a complex at the site of the axon-glial junction and play a part in the initiation of myelination. Consistently, we found that Par-3 and $p75^{NTR}$ are colocalized at the axon-glial junction (Fig. 3A). In addition, Par-3 coimmunoprecipitates with an antibody to $p75^{NTR}$, and vice versa, in newborn mouse sciatic nerve extracts, which suggests an interaction in vivo (Fig. 3B). To investigate the timing of the Par-3/$p75^{NTR}$ association, we turned to the SC/DRG coculture system, where we can precisely control the onset of myelination. As expected, Par-3 and $p75^{NTR}$ were detected in the cocultures both before and after the induction of myelination, but 10 days after induction, when the large majority of myelin had already formed, Par-3 and $p75^{NTR}$ protein levels were reduced (fig. S6), suggesting that they may be required only in the early stages of myelination. To test this hypothesis, we coimmunoprecipitated Par-3 with an antibody to $p75^{NTR}$, using protein extracts isolated at different time points from the SC/DRG cocultures (Fig. 3C). We found that Par-3 and $p75^{NTR}$ did not interact in SC/DRG cocultures before the induction of myelination, whereas after induction, Par-3 and $p75^{NTR}$ transiently associated, peaking at 2 days after induction (Fig. 3C). This finding was further confirmed by quantifying the percentage of SCs that displayed either asymmetric Par-3 or both asymmetric Par-3 and $p75^{NTR}$ by immunostaining. About 10% of the total number of SCs displayed asymmetric Par-3, but only 5% of the SCs displayed a colocalization of Par-3 and $p75^{NTR}$ (Fig. 3D). This colocalization was transient and peaked at the initiation of myelination (Fig. 3D).

Because BDNF is known to promote SC myelination by acting through $p75^{NTR}$, and the removal of BDNF inhibits this process, we sought to determine the specific nature of the Par-3/$p75^{NTR}$ interaction in the myelination program, which would be expected to increase in conditions favorable for myelination. Consistent with this possibility, the addition of exogenous BDNF to the culture medium enhanced the Par-3/$p75^{NTR}$ association beyond that of control cultures, and removing endogenous BDNF, using the BDNF scavenger TrkB-Fc, significantly inhibited this association (Fig. 3E). These results were recapitulated by examining the colocalization of Par-3 and $p75^{NTR}$ by immunostaining in the presence or absence of BDNF. The percentage of cells that displayed both asymmetric Par-3 and $p75^{NTR}$ doubled with the addition of exogenous BDNF. These results demonstrate a ligand-dependent association of Par-3 and $p75^{NTR}$ and provide further evidence for a specific interaction and function that is critical for myelination.

The $p75^{NTR}$ has a PDZ binding motif at the C terminus (*20*), and Par-3 contains three PDZ domains (*11*), raising the possibility that Par-3 and $p75^{NTR}$ might associate via PDZ domains. To test this hypothesis, we constructed retroviral vectors encoding both GFP and each of the three PDZ domains of Par-3, or a 20 amino acid portion of the intracellular domain (ICD) of $p75^{NTR}$, which contains the PDZ binding motif. Infection of SCs cocultured with DRG neurons with each of the constructs revealed that overexpression of $p75^{NTR}$-ICD and the PDZ-1 and PDZ-2 domains of Par-3 significantly inhibited myelination, but overexpression of the PDZ-3 domain of Par-3 had no effect (Fig. 4, A and B). To determine whether Par-3 directly interacts with $p75^{NTR}$, we used an in vitro binding assay with purified FLAG-tagged PDZ domains and a glutathione *S*-transferase (GST)–$p75^{NTR}$-ICD. Remarkably, we found that only the PDZ-1 domain of Par-3 binds the ICD of $p75^{NTR}$, whereas the other PDZ domains did not associate with $p75^{NTR}$-ICD (Fig. 4C).

Is Par-3 required for the asymmetric localization of $p75^{NTR}$, or vice versa? To address this question, we counted the proportion of SCs that displayed asymmetric distribution of $p75^{NTR}$ after Par-3 knockdown, compared with control-infected SCs. The number of SCs that displayed asymmetric $p75^{NTR}$ localization after Par-3 knockdown (19.2 ± 3.9% with control shRNA versus 8.1 ± 0.8% with Par-3 shRNA) was reduced by half, indicating that the asymmetric localization of $p75^{NTR}$ requires Par-3. Although some SCs still display asymmetric $p75^{NTR}$ after Par-3 knockdown, it is likely that this is due to residual Par-3 expression after shRNA knockdown. To determine whether $p75^{NTR}$ could be required for the asymmetric localization of Par-3 at the axon-glia junction, we analyzed cocultures of SCs and DRG neurons prepared from $p75^{NTR}$

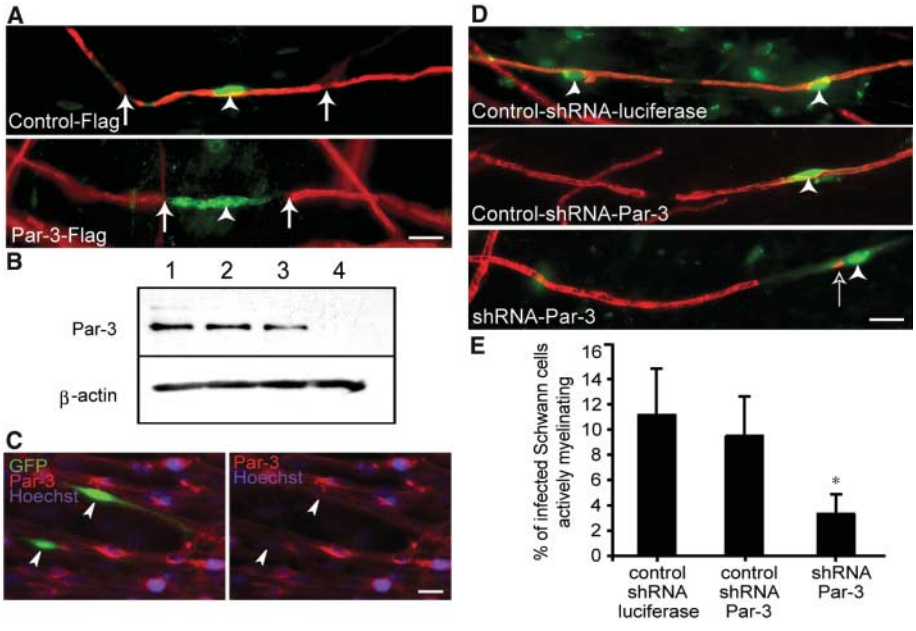


Fig. 2. Disruption of Par-3 localization and function by overexpression or knockdown inhibits SC myelination. (**A**) Control-infected SCs (top; green, arrowhead) formed myelin normally as detected by immunostaining for MBP (red). Arrows indicate the length of the myelin internode. Par-3–overexpressing cells (bottom; green, arrowhead) fail to form myelin. (**B**) The efficiency of the shRNA for Par-3 knockdown was examined in purified rat SCs. Cells were either not transfected (lane 1) or were transfected with a retroviral vector expressing GFP only (lane 2), a nonblocking shRNA for Par-3 (lane 3), or the experimental shRNA for Par-3 (lane 4); Par-3 protein expression was detected by Western blot. (**C**) The Par-3 shRNA retroviral vector was also tested by infecting SCs in cocultures, and 3 days later Par-3 expression was detected by immunostaining (red). Infected cells (green, arrowheads) show a dramatic reduction of Par-3 expression compared with adjacent noninfected cells. (**D**) SCs cocultured with DRG neurons were infected with either control or Par-3 shRNA retroviral vectors. Infected cells express GFP (green, arrowheads), and myelin is detected by staining with MBP (red). (**E**) Quantification of the Par-3 knockdown experiment was achieved by counting the percentage of infected SCs actively myelinating. Error bars, SD; $*P < 0.005$, unpaired Student's *t* test. Scale bars, 10 μm.

knockout mice (*21*). We found that p75^{NTR} knockout SCs displayed the same asymmetric Par-3 localization as wild-type SCs (fig. S7), indicating that asymmetric localization of Par-3 does not require p75^{NTR}.

The results presented here identify a novel protein-protein interaction via the PDZ binding motif of p75^{NTR}, to our knowledge the first demonstration of Par-3 directly associating with a membrane receptor. Our data also show an essential requirement for the establishment of SC polarity in myelination, because disruption of Par-3 localization by overexpression and knockdown, which in turn disrupts the Par-3/p75^{NTR} complex, inhibits SC myelination. Because overexpression of Par-6, another polarity protein known to form a complex with Par-3, can also inhibit myelination (fig. S8), we propose that the Par polarity complex has a major role in SC myelination. Based on our results, we suggest the following model (Fig. 4D). SCs become polar-

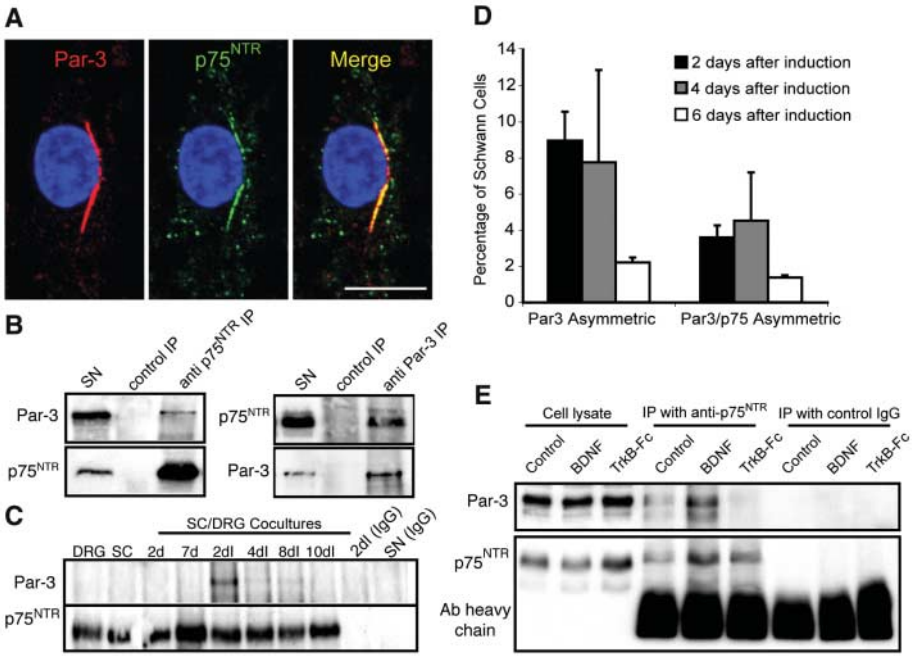


Fig. 3. Par-3 and p75^{NTR} associate transiently at the initiation of myelination. (**A**) Premyelinating SCs cocultured with DRG neurons displayed asymmetric localization of Par-3 and p75^{NTR}. Colocalization was observed (yellow) when the corresponding images were merged. A z-stack of images was acquired with structured illumination fluorescence microscopy and reconstructed in three dimensions to confirm colocalization. A projection of the z-stack is shown. Scale bar, 10 μm. (**B**) Coimmunoprecipitation for Par-3 and p75^{NTR} was performed using sciatic nerves (SN) of newborn mouse pups. Par-3 coimmunoprecipitated with an antibody to p75^{NTR} and was detected by immunoblotting for Par-3, whereas p75^{NTR} coimmunoprecipitated with an antibody to Par-3 and was detected by immunoblotting for p75^{NTR}. Normal rabbit immunoglobulin Gs were used as controls for the immunoprecipitation. (**C**) Purified DRG neurons, SCs, and cocultures at different days in culture (d) and days after induction of myelination (dI) were used for the coimmunoprecipitation of Par-3 with an antibody to p75^{NTR}. Par-3 was detected by immunoblotting for Par-3 (top). The blot was then stripped and reprobed with an antibody to p75^{NTR} (bottom). (**D**) Quantification SCs displaying only asymmetric Par-3 or both asymmetric Par-3 and p75^{NTR} in SC/DRG cocultures. The cultures were fixed and stained at 2, 4, or 6 days after induction of myelination. Error bars, SD. (**E**) Coimmunoprecipitation with the p75 antibody from cocultures at 2 days after induction of myelination in the presence of exogenous BDNF, or the BDNF scavenger TrkB-Fc. The blot was probed for Par-3 (top), stripped, and reprobed for p75^{NTR} (bottom).

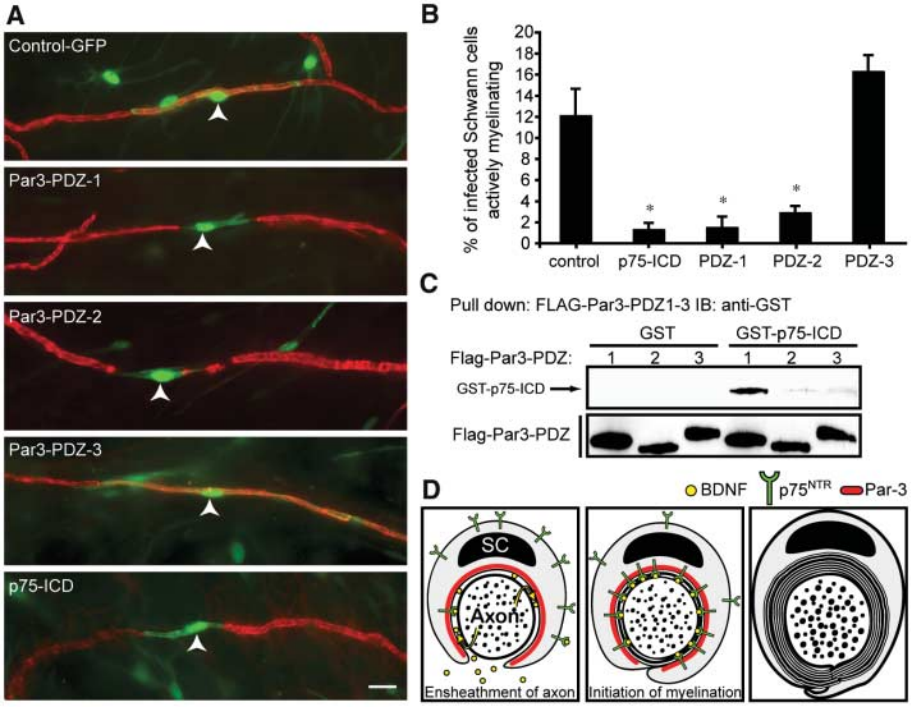


Fig. 4. Par-3 directly associates with p75^{NTR} via its PDZ1 domain. (**A**) Retroviral vectors expressing GFP, and each of the three PDZ domains of Par-3, or the ICD of p75^{NTR}, were generated and used to infect SCs cocultured with DRG neurons. Infected cells overexpressed the individual domains of Par-3 and p75^{NTR} and GFP (green, arrowheads). Myelin was identified by immunostaining for MBP (red). Scale bar, 10 μm. (**B**) Quantification of the overexpression experiment was achieved by counting the proportion of infected SCs actively myelinating. Error bars, SD; *$P < 0.001$, unpaired Student's t test. (**C**) The FLAG-tagged version of the PDZ domains of Par-3 and the bacterially produced GST-p75-ICD proteins were purified, and binding was assayed in a cell-free extract. The complex was pulled down by immobilizing the FLAG-PDZ domains, and detection of binding was determined by immunoblotting with an antibody to GST. (**D**) Schematic representation of a proposed model for the role of Par-3 in SC myelination. An illustration of the cross section of a myelin internode at different time points throughout the myelination process.

ized upon contact with axons, as reflected by the asymmetric enrichment of Par-3 at the site of the axon-glial junction (Fig. 4D, left). As SCs prepare to initiate myelination, BDNF is synthesized by DRG neurons, secreted and released along axons to promote SC myelination. BDNF binds to $p75^{NTR}$ on the SCs, which somehow accommodates the association between the PDZ-1 domain of Par-3 and the PDZ binding motif of $p75^{NTR}$ (Fig. 4D, left). Concomitantly, Par-3 recruits $p75^{NTR}$ to the axon-glial junction, ensuring that BDNF is continuously and efficiently received by the SC and translated into signaling events that lead to the formation of myelin (Fig. 4D, left and middle). Once myelination is complete, Par-3 and $p75^{NTR}$ are down-regulated, and the complex dissociates (Fig. 4D, right).

Our data raise the possibility that the asymmetric localization of Par-3 at the axon-glial junction causes a convergence of various receptors and adhesion molecules that are essential for the initiation of myelin formation. This is supported by our observation that overexpression of the PDZ-2 domain, which does not bind $p75^{NTR}$-ICD, also efficiently abolishes myelination by SCs. It will be interesting to characterize the molecules that are recruited to this site and to determine their role in the process of myelination. Additionally, it will be important to determine whether cell polarity mechanisms are involved in nerve regeneration and in remyelination paradigms, because this will largely impact the design of therapies for various peripheral neuropathies and nerve injury.

References and Notes

1. G. V. Michailov *et al.*, *Science* **304**, 700 (2004).
2. J. R. Chan, J. M. Cosgaya, Y. J. Wu, E. M. Shooter, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14661 (2001).
3. J. M. Cosgaya, J. R. Chan, E. M. Shooter, *Science* **298**, 1245 (2002).
4. J. Yamauchi, J. R. Chan, E. M. Shooter, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14421 (2003).
5. J. Yamauchi, J. R. Chan, E. M. Shooter, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8774 (2004).
6. E. M. Pinheiro, D. J. Montell, *Development* **131**, 5243 (2004).
7. M. Schober, M. Schaefer, J. A. Knoblich, *Nature* **402**, 548 (1999).
8. S. H. Shi, L. Y. Jan, Y. N. Jan, *Cell* **112**, 63 (2003).
9. A. Wodarz, A. Ramrath, U. Kuchinke, E. Knust, *Nature* **402**, 544 (1999).
10. K. Ebnet *et al.*, *EMBO J.* **20**, 3738 (2001).
11. B. Etemad-Moghadam, S. Guo, K. J. Kemphues, *Cell* **83**, 743 (1995).
12. Y. Izumi *et al.*, *J. Cell Biol.* **143**, 95 (1998).
13. G. Joberty, C. Petersen, L. Gao, I. G. Macara, *Nat. Cell Biol.* **2**, 531 (2000).
14. D. Lin *et al.*, *Nat. Cell Biol.* **2**, 540 (2000).
15. S. Ohno, *Curr. Opin. Cell Biol.* **13**, 641 (2001).
16. I. B. Wanner, P. M. Wood, *J. Neurosci.* **22**, 4066 (2002).
17. T. J. Hung, K. J. Kemphues, *Development* **126**, 127 (1999).
18. J. Nance, J. R. Priess, *Development* **129**, 387 (2002).
19. H. Zhang, I. G. Macara, *Nat. Cell Biol.* **8**, 227 (2006).
20. P. P. Roux, P. A. Barker, *Prog. Neurobiol.* **67**, 203 (2002).
21. D. von Schack *et al.*, *Nat. Neurosci.* **4**, 977 (2001).

22. We dedicate this manuscript to E. M. Shooter for his endless support, encouragement, and wisdom. We thank P. Barker for the p75 knockout mice, I. Macara for the Par-3 shRNA constructs, T. Pawson for the Par-3 antibody, D. Meijer for the Krox-20 antibody, and A. Vallée of the IRCM histology core facility for expert technical assistance with electron microscopy. We are grateful to W. Mandemakers and A. Kania for reviewing this manuscript. This work was supported by the National Multiple Sclerosis Society Career Transition Award (J.R.C.) and the Canadian Institutes of Health Research (M.C).

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/832/DC1
Materials and Methods
Figs. S1 to S8
References

18 August 2006; accepted 26 September 2006
10.1126/science.1134069

Autosampler

The Vortex autosampler consists of an integrated autosampling device and multi-stirrer. It is designed to increase efficiency, enhance precision, and promote operator safety. The temperature-controlled, heated (up to 150° C) multi-stirrer assembly features a stainless steel tray, capable of handling both 20 ml and 40 ml vials. Each vial position has its own magnetic stirring mechanism, delivering consistent and reproducible stirring action from vial to vial and eliminating variability due to the dissolution process. When coupled with Viscotek's Triple Array Detector or Relative Viscometer, the Vortex provides a complete, automated system for macromolecular characterization and determination of solution viscosities.

Viscotek For information 800-375-5966 www.viscotek.com

Protease Assay Kit

The Protease Assay Kit can be used in conjunction with the Protease Inhibitor Set to determine which proteases are present in a sample. The Protease Assay Kit makes use of a dye-labeled protein conjugate as a protease substrate that allows for the detection of nanogram levels of protease activity. A sequential analysis of the sample using the 11-protease inhibitor set will quickly determine the active proteases. The results allow the researcher to determine which protease inhibitor combinations are most advantageous and their optimal concentrations.

G-Biosciences/Genotech For information 800-628-7730 www.GBiosciences.com

Protein Microarray for Antibody Characterization

UNIchip protein microarrays generate a quantitative binding profile of an antibody to 400 different, unrelated targets. Together with on-chip determination of sensitivity, linearity, and dynamic range, the quantitative fingerprint enables a unique performance ranking of antibodies. Designed to save time on downstream assays by enabling early selection of the best antibodies for validation, the protein biochips are available with either a cross section of 400 human proteins or specific protein sets.

Protagen For information +49 231 9742-6300 www.protagen.com

Dissection Tools

The Harris Micro-Punch and Uni-Cores are designed to serve research scientists with a dissection tool for small predefined regions of tissue or specimens. Uni-Cores are disposable, multi-purpose sampling tools for soft substrates. Applications include, in electrophysiology, punch dissection of discrete regions prior to acute dissociation for patch recording; in biomedical pharmacology, punch removal of discrete brain regions for analysis of neurotransmitters and metabolite changes in response to different pharmacological agents; in anatomy, punch-out of small brain regions for analysis of neurotransmitter concentrations of messenger RNA levels or for subsequent use for electron microscopy; positioning samples inside quartz pyrolysis gel tubes; and positioning samples on Fourier transform infrared diamond cell sample stages.

Electron Microscopy Sciences For information 215-412-8400 www.emsdiasum.com

Recombinant Baculoviruses

BacMagic DNA provides selection of baculovirus recombinants for protein expression in insect cells, eliminating the time-consuming plaque purification steps. BacMagic DNA is an AcNPV genome with a deletion in the essential open reading frame (ORF) 1629 and an added bacterial artificial chromosome (BAC). This combination prevents nonrecombinant virus from replicating in insect cells and allows the viral DNA to be propagated in bacterial cells. A compatible transfer plasmid containing the target protein coding sequence is cotransfected with BacMagic DNA into insect cells.

EMD Biosciences For information 800-628-8470 www.novagen.com

Literature

Protein Folding is an application note demonstrating that the DynaPor Dynamic Light Scattering (DLS) system is an effective alternative to in vitro experiments for following the protein unfolding process. Protein folding and unfolding have historically been monitored via in vitro experiments, in which the information obtained is directly related to the function and behavior of proteins in vivo. The aim of this experiment was to follow the unfolding process by another method, dynamic light scattering. The DynaPro system offers high sensitivity and low sample volumes for protein crystallization by DLS. Features include a control module with the ability to collect and analyze data and a read module, which comes as either a temperature-controlled scintillation vial reader for single measurements or a fully automated plate reader with the capability to make measurements on 96-well, 384-well, or 1536-well plates. In this application, the DLS measurements were found to be capable of monitoring the dissociation of an oligomeric protein and unfolding of monomers simultaneously using chaotropic agents. The application note can be downloaded from the web address below.

Wyatt Technology For information 805-681-9009 www.wyatt.com/literature

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier by visiting www.science.labvelocity.com on the Web, where you can request that the information be sent to you by e-mail, fax, mail, or telephone.

NEWS

U.S. Life Scientists Report Rising Salaries and High Job Satisfaction

The latest salary survey of AAAS life scientists, developed in partnership with Kelly Scientific Resources, finds that pay increases outpaced inflation, especially for postdocs

Becoming a scientist isn't easy. Earning a Ph.D. is just one step in the long journey to professional fulfillment. Then comes a postdoc or three, and maybe a multiyear probationary period. Most scientists are in their mid-30s by the time their real careers get under way. And the payoff—the opportunity to spend the rest of your life working really hard on interesting projects—brings with it a decent, but usually not great, salary.

Fortunately for the profession, most scientists seem willing to put up with those challenges. Those who responded to the third salary and employment survey by AAAS (publisher of *Science*) report job satisfaction at just a shade below "very good." "I can't believe I'm getting paid for this!" says Bonny Dickinson, an assistant professor at the Research Hospital for Children at the Louisiana State University Health Sciences Center in New Orleans. And in a year when most workers saw their real incomes fall or stagnate, life scientists report receiving modest but significant salary gains. "I think my compensation is good," says Linval DePass, executive director of nonclinical safety at Durect Corp., a Cupertino, California, specialty pharmaceuticals company. "I think salaries are increasing, particularly in the area where I work."

For life scientists, 2005 was a pretty good year. Full-time academic life scientists—the largest group in our survey—reported earning 5.4% more this year, on average, than they did the previous year. That's well above the 3.4% rate of inflation. Full-time scientists in industry did quite a bit better, earning a boost of nearly 10%. To no one's surprise, the salaries of industrial scientists with doctorate degrees far outpace those of their academic colleagues, with means of $116,000 and $78,000 respectively. That difference, coincidentally, nearly matches the $40,000 average salary being collected by academic postdocs in 2006.

Still, even those academic workhorses are doing a little better than in the past. Salaries for postdocs—17% of respondents—rose by an average of 8.1%, more than double the inflation rate. Postdocs in industry had a blockbuster year, earning 19% more. (One caveat to the postdoc numbers: Some of the respondents may be first-year postdocs, meaning the comparison may be to what they were earning as graduate students.)

Despite the rising tide, large disparities remain. Even excluding industrial scientists, scientists in certain disciplines earned far less than their colleagues did. Developmental biologists with doctoral degrees who work in academia, for example, earned a median salary of just $45,000, whereas Ph.D. pharmacologists earned $99,000. Academic developmental biologists gained some ground, however, earning 7% more than last year, compared to a 2% average raise for Ph.D. toxicologists.

How the survey was conducted

The survey targeted U.S.-based life scientists. We sent e-mail invitations to 41,000 AAAS members and 12,000 free registrants on the *Science* magazine Web site. Kelly Scientific Resources

CREDITS: JUPITER IMAGES

also participated in the survey by polling some 12,000 of their employees, whose responses were combined with the rest of the survey data.

The overall response rate was 7%—just over 4500 scientists—of whom 62% are employed in academia. Some 35% work in industry, government, or the nonprofit sector, and 3% are self-employed. Some 43% of survey respondents are women, three-quarters hold doctoral-level degrees, and 22% are nonwhite. Nearly 9 in 10 are U.S. citizens or permanent residents.

The information collected goes well beyond their paychecks. They were asked where they work and for how long, what they do, and how satisfied they are. (Of the 12% working on temporary visas, for example, half report that their immigration status has caused "job security/pay issues," and 23% said that they faced "challenges obtaining grants.") We also invited them to make comments and interviewed several people who agreed to be quoted.

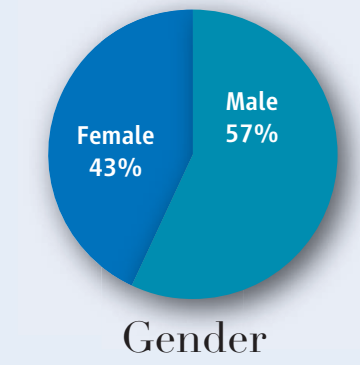
Despite the considerable challenges of a scientific career—relatively low pay, long hours, a very long training phase, and a tight job market—scientists say they enjoy the work. Survey respondents rated their job satisfaction as 3.7 on a 5-point scale, between "good" and "very good." The finding is similar to those of the two previous surveys (*Science*, 18 June 2004, p. 1829, and 12 October 2001, p. 395), suggesting that scientists' happiness is not a new phenomenon.

You might think that postdocs, with their low salaries, poor job security, and often-poor working conditions, would be less happy than most other scientists. They were—but not by much. Postdocs rated their job satisfaction at 3.5, midway between "good" and "very good."

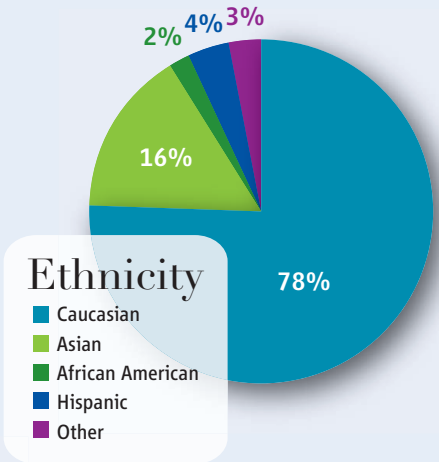
So what makes scientists happy? Surely it's not simply how much they're paid; compared to the earnings of those in other professions requiring similar training, scientists' salaries remain quite low. But our survey showed that salary *is* one of many factors determining job satisfaction. People who rated their salary "excellent"—a 5—were also more than three times as likely as the average survey respondent to rate their overall job satisfaction as "excellent."

But a closer look at the connection between salary and job satisfaction shows that the correlation really isn't all that strong. A linear regression analysis of job-satisfaction versus salary shows that whereas scientists who earned $150,000 rated their job satisfaction as "very good," scientists earning barely one-fifth as much were only slightly less satisfied. Furthermore, scientists who

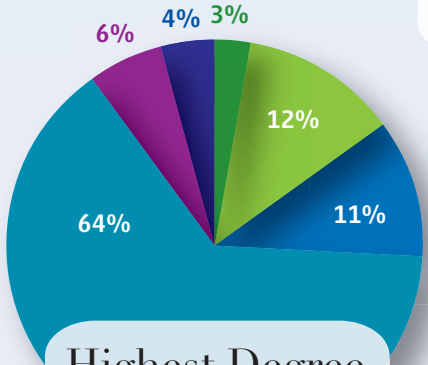
Survey Demographics



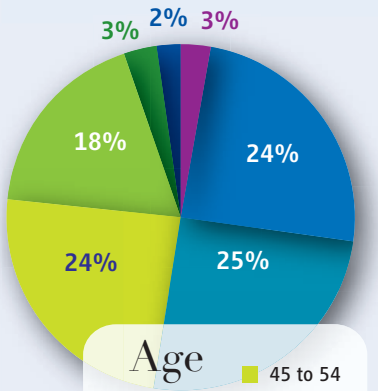
Gender



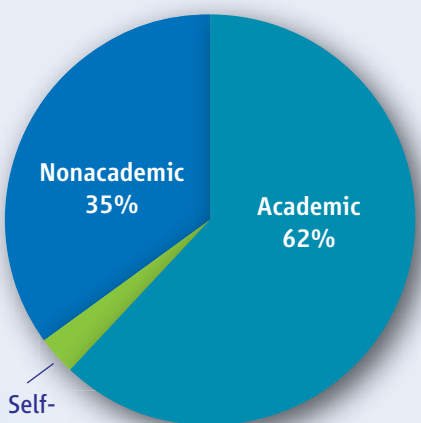
Ethnicity
- Caucasian
- Asian
- African American
- Hispanic
- Other



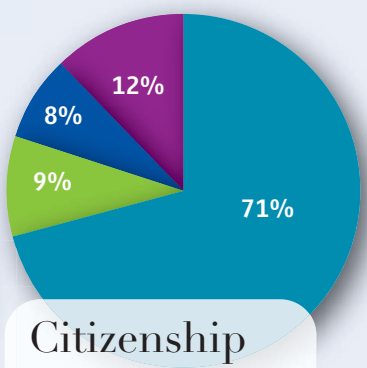
Highest Degree Received
- Ph.D.
- Master's
- Bachelor's
- M.D.
- M.D./Ph.D.
- Other



Age
- Under 25
- 25 to 34
- 35 to 44
- 45 to 54
- 55 to 64
- 65 to 70
- Over 70



Type of Organization

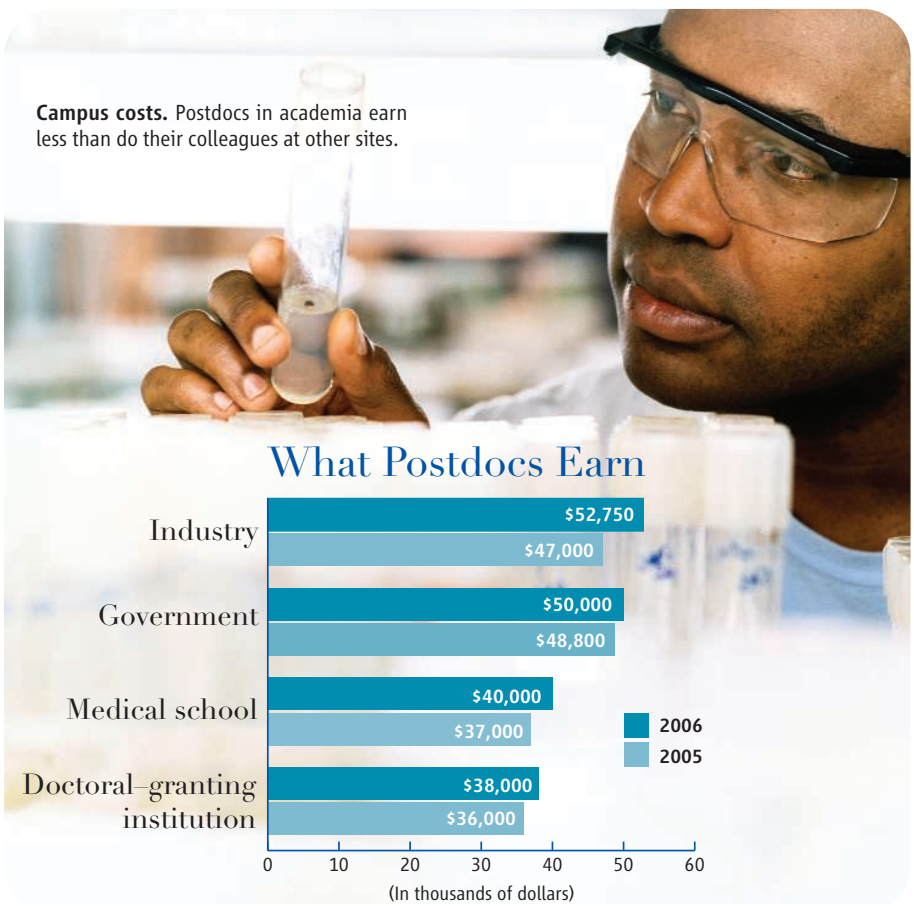


Citizenship
- U.S. native citizen
- U.S. naturalized citizen
- U.S. permanent resident visa
- Other temporary visa

Look who's talking. Survey respondents are well-distributed by age and most likely to be white male Ph.D.s working at an academic institution.

Median 2006 Salaries

	Academic	Industry
Medicine	$140,000	$145,500
Pharmacology	$99,000	$116,000
Toxicology	$78,000	$105,000
Environmental	$85,000	$88,000
Genetics	$73,458	$83,750
Agricultural	$74,000	$99,500
Biochemistry	$63,300	$98,000
Physiology	$65,000	$88,010
Neuroscience	$66,150	$100,000
Biotechnology	$63,000	$105,000
Microbiology	$62,000	$89,000
Bioinformatics	$65,000	$103,000
Ecology	$61,474	$75,000
Zoology	$59,300	$47,000
Virology	$57,500	$89,000
Molecular biology	$52,750	$84,000
Immunology	$54,000	$96,525
Cell biology	$52,050	$84,300
Cancer biology	$50,000	$62,000
Developmental biology	$45,000	$65,130
Other	$71,000	$92,000



Campus costs. Postdocs in academia earn less than do their colleagues at other sites.

CREDIT: GETTY IMAGES

work in industry report exactly the same level of job satisfaction as their lower-paid colleagues in academia.

What else affects scientists' job satisfaction? Among the most important factors are promotion opportunities, job security, and intellectual challenge. Our follow-up interviews identified still other factors and gave a more nuanced view of scientists' deepest professional desires.

Moving ahead

Traditionally, Harvard University has been notoriously stingy with tenure. But neuroscientist Florian Engert says times have changed, at least in his corner of the Yard. "I can think of five people who have gotten tenure in the last few years," he says, "and only one who hasn't." (A Harvard administrator told *Science*, "We have not released tenure numbers in the past and would not do so at this time.") Accordingly, Engert, who comes up for tenure in 2 years, rates his job satisfaction as "excellent."

Engert, an associate professor in Harvard's Department of Molecular and Cellular Biology, studies vision and cognition by monitoring brain activity in zebrafish. Engert's "excellent" job-satisfaction rating is consistent with the fact that people who rated their promotion opportunities as "excellent" were four times more likely than average to rate their job satisfaction as "excellent," too. Likewise, immunologist Kathleen Hoag of Michigan State University in East Lansing loves her job despite the long hours—"It's fun! Intellectually challenging! Tiring!"—and rates her chances of getting tenure this year as "almost certain."

The correlation is even stronger in the other direction. Those who rate their promotion opportunities "very poor" are seven times more likely than average to register low job satisfaction. Many postdocs and soft-money researchers find themselves in precisely that position.

"To start a scientific career is very challenging, and the returns are not very good," says Rahul Sharma, a research instructor at the University of Virginia in Charlottesville, who sees his promotion opportunities and job satisfaction as very poor. (The university uses the title "research instructor" for those who have exceeded the stipulated 5-year stint as a postdoc.) "It's more than 7 years since I got my Ph.D., and I'm still struggling to start my career. This is a critical time."

Sharma hopes to snare a soft-money research-faculty position soon. "These [jobs] are a little better than being a post-

doc," he says. Besides paying him better, such a job would give him the chance to write his own grant proposals and leverage any funding into a position with greater stability. "Until you are a tenured faculty member, the career doesn't really start," says Sharma, who plans to reassess his situation in a couple of years if his prospects haven't improved.

The right fit

Whereas job security boosts professional satisfaction, our survey found, there's no sure path to fulfillment. Ken Nussbaum, a veterinary epidemiologist at Auburn University in Alabama, sees a "burgeoning need" for animal experts in public health, and he wants to help fill that need. Nussbaum has tenure. But he gives himself a 1 for job satisfaction, and he's in the job market.

"I seem to have plateaued academically," Nussbaum says. "I don't wish to continue highly technical lab work, yet I have training that I hope will be transferable to the field of public health. I'd like to spend a couple of years working in public health or step back into more concentrated classroom teaching." Although his efforts to enter the public-health field are going well—he has had a paper accepted at an important meeting—he hasn't yet succeeded in finding a new job.

"I think salaries are increasing, particularly in the area where I work."
◀ Linval DePass

"In some places, it's a [professional] mismatch," he says. Despite his 25 years' experience, employers are offering him entry-level positions. "At my stage, my strength is my understanding," he says.

Nussbaum has found few opportunities close to home. Although he is willing to relocate, he has ruled out some cities. "A lot of jobs in the Washington, D.C., area look like good positions," he says, but he's scared off by the higher cost of living, in particular the disparity with the deep South on housing.

Steve Verhey has also struggled to mesh his work with his values and ambitions. When he completed the survey, Verhey was a self-described "lame-duck" assistant professor of cell biology at Central Washington University in Ellensburg, where he had just been denied tenure. His frustration at not being recognized for focusing on teaching at the predominantly undergraduate institution was reflected in his self-rating of 1 for job satisfaction.

But concentrating on research wasn't a viable option, either, he says, because of the lack of resources—such as lab space—and a balky infrastructure. "The thing that got me the most was the purchasing system," he says. "Suppliers are set up so that you can call, and they'll ship it the next day." But at Central Washington, he says, orders might languish for a week. "Not conducive to doing real research," he says.

So Verhey went out and founded his own biodiesel company. In the process, he raised his job satisfaction to a 5. "I'm having more fun that I ever did in 20 years in academia," he says. "I'm using things I've learned and learning new things—in the real world."

Although some people may be happy to labor in obscurity, our survey identified "recognition and prestige" as one of the most important factors in determining job satisfaction. People who gave the top mark to the "recognition and prestige" of their jobs were 3.5 times more likely than average to rate their overall job satisfaction just as high. On the other end, those who gave "recognition and prestige" a 1 were four times more likely than average to give their overall job satisfaction the same poor rating.

In interviews, individual scientists are loath to cite recognition and prestige as keys to professional happiness. But Peter Koulen, a tenured full professor at the University of North Texas Health Science Center in Fort Worth, readily admits that prestige—including institutional prestige—has important practical consequences. "A lot of professional success critically depends on the name of your institution, which, if peer review works the way we think it works,

"She's been like a cheerleader for me, understanding that I have a life outside the lab."
Kathryn Shows ▶

Mind the gap. Women gained slightly on men last year, but their salaries still lag.

CREDITS (TOP TO BOTTOM): COURTESY OF L. DEPASS; COURTESY OF K. SHOWS; GETTY IMAGES

should not be the case," says Koulen, who has also worked at Yale University and at the Marine Biological Laboratory in Woods Hole, Massachusetts. "I think the quality of my science has not changed, but the work that I have to do to get my science published or funded is multiple times harder here."

Carol Sibley, a full professor in the recently created genome sciences department at the University of Washington, Seattle, says she enjoys "excellent" job satisfaction but is considering a change nonetheless. "I'm 62, [and I] still have lots of energy and interest. But at some point you think, 'I've been a prof all this time; maybe there are other ways to apply my skills.'"

Sibley, who studies the two malaria parasites, would like her work to have more direct relevance to public health. "I study drug resistance. As you work on that at the basic-science level, you get more and more interested in how we can slow this [disease] down." Living in the city that's home to the Bill and Melinda Gates Foundation and other organizations focused on world health, she acknowledges that she has many opportunities. But she wants one with modest travel requirements. "One of the minuses is that the folks I know travel a huge amount—more than I want to," she says.

"I can think of five people who have gotten tenure [at Harvard] in the last few years and only one who hasn't."

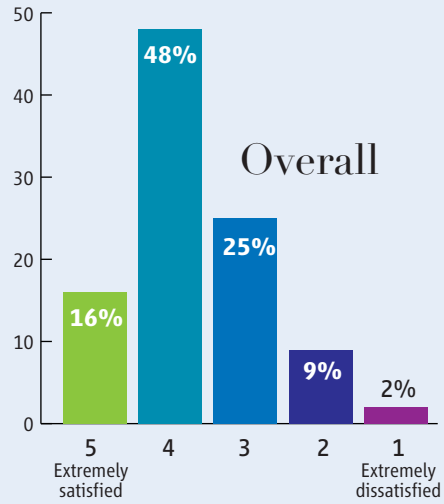
Florian Engert ▶

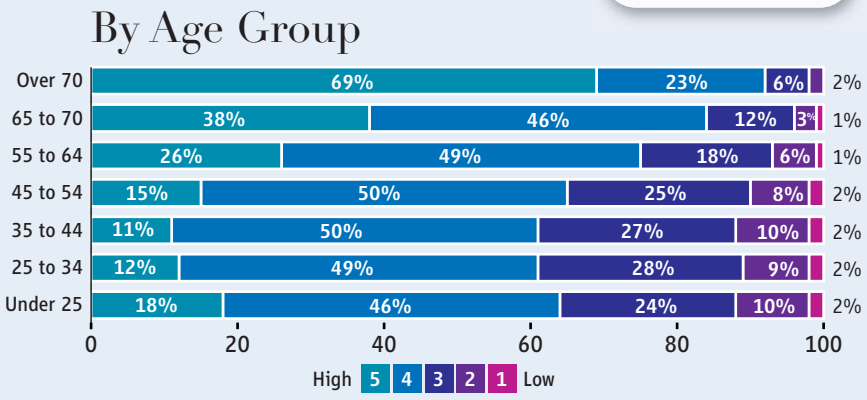
The right location

Geography, too, can have a big impact on scientists' job satisfaction. Scientists who rated their geographic location "not satisfactory" were three-and-a-half times more likely than other respondents to suffer from low job satisfaction. Christopher Dougherty, a soft-money research professor at Florida Atlantic University (FAU) in Boca Raton, studies the genetics of age-related macular degeneration. His goal is something more remunerative and secure. But he and his wife prefer to stay in town so their 6-month-old son can get to know his grandparents, who live nearby.

Dougherty's solution is to wait for the jobs to come to him. Scripps Research Institute has just committed to opening a new research facility—with a research agenda that matches up well with Dougherty's focus on the genetics of geriatric diseases—on FAU's other campus in nearby Jupiter. Scripps is expected to attract a cluster of new biotech companies to the area.

Job Satisfaction





Working it out. A good fit, including the chance to move up in the right setting, plays a huge role in determining job satisfaction for scientists of all ages—although older workers report being especially content.

CREDIT: COURTESY OF F. ENGERT

But just as some scientists are limited by geography, others are aided by it. "I've worked in the [San Francisco] Bay area, which is a very strong area for biotech," for 22 years, says Durect's DePass. "Because of that, there is considerable competition for good people, which contributes to an increase in salaries."

The satisfaction of older scientists

Conventional wisdom and at least a few scholarly articles suggest that scientists do their most creative work when they're still young. Our survey indicates, however, that it's older scientists who find their work most satisfying. The upward trend in job satisfaction begins at 35, but our survey found the biggest jump at 55.

So why are older scientists happier? One reason may be that they get paid a lot more than their junior colleagues do. Everyone knows that older workers earn more than younger ones, but our survey showed that in the life sciences the trend is stronger and continues longer than in the general population. Academic life scientists above 65 reported mean salaries of $133,000, more than three times as large as those in the youngest (25 to 34) age group, who earn a mean of $41,000. For the general population, the ratio of peak earnings to early-career earnings is only 1.24, according to the Bureau of Labor Statistics. BLS data also show that workers in the general population achieve peak earnings between 45 and 54, whereas in our survey, the very oldest science workers also report the very highest salaries.

So although young life scientists earn only a little more than the typical American worker, their salaries increase much more steeply, and the increases persist. That pattern makes sense to David Inouye, a 56-year-old ecologist and conservation biologist at the University of Maryland, College Park. He says his job satisfaction "stems in part from the success of a graduate program that I started 18 years ago." On the research side, the full payoff was also slow in coming.

"In terms of research, much of what I do now builds upon work that I started as a graduate student in 1973," he says. "So part of the age-satisfaction correlation, at least in my own case, comes from having laid the groundwork decades ago for work that is still ongoing and being very fruitful now."

For Ananda Lugade of Luminex Corp. in Austin, Texas, the correlation has a simpler explanation: She just didn't find the right job until she was older. "I could not get satisfaction at an early age," she wrote in an e-mail.

By Age

Age	Salary
25 to 34	$41,000
35 to 44	$62,000
45 to 54	$87,000
55 to 64	$112,000
65 to 70	$133,000
Over 70	$133,000

Golden years. Scientists' salaries keep rising throughout their careers.

"I got an opportunity to make some good contributions," she says—but only after she reached 45.

Mentorship and guidance

So what about younger scientists? Our survey did not address directly the issue of mentorship, but almost all the young scientists we interviewed call this an important factor. Michigan State's Hoag is grateful that her department has faculty mentoring committees, which are available to advise all probationary faculty members. Hoag's department offers all faculty members an annual review, which allowed Hoag to be clear on what was expected of her and how well she was meeting those expectations. That knowledge has been an important element of her professional contentment.

Kathryn Shows, a postdoc at Virginia Commonwealth University in Richmond who studies the genetic disorder Treacher Collins syndrome, credits her postdoctoral adviser and mentor for helping her return to science after leaving to start a family: "She encouraged me to come back, and she's been like a cheerleader for me, understanding that I have a life outside the lab." Shows also feels she's getting the kind of training she needs to develop into an independent academic scientist. "I've always been interested in genetics and cannot imagine studying anything else. I [also] cannot imagine being anywhere else than at a university."

"It's more than 7 years since I got my Ph.D., and I'm still struggling to start my career."

▼ Rahul Sharma

CREDIT: COURTESY OF R. SHARMA

Dickinson—the Louisiana State immunologist who said she couldn't believe she's getting paid for the interesting work she's doing on the mechanism of cholera, among other things—credits her mentor, Seth Pincus, for making a big difference in her professional life. "I'm happy to be a part of a team working under his leadership," she says. She especially admires how Pincus got the lab back up and running 6 weeks after Hurricane Katrina devastated New Orleans.

Despite her overall satisfaction, she admits the fit isn't perfect. Her institution isn't large, and compared to Harvard, where she did a postdoc, its approach to science is a bit low-key, she says. "It's a little bit isolating," she confesses. "I gave a seminar yesterday, but there were so few people there. I spent so much time preparing, and nobody gave me any feedback. Sometimes I really miss Harvard, where people are very excited. It's not like that here; come 5 o'clock, it's hard to find people."

Salary matters for Dickinson. But what matters most is the peace of mind it buys. "I make $84K," Dickinson says. "I would still do it at $65K, [but] I'd be spending more time worrying about finances and so on." In the meantime, she says, her life is good. "My experiments are going well. I feel blessed."

Another factor our survey didn't address directly—but that came up repeatedly in comments and interviews—is the importance of good colleagues. And Michigan State, with the sixth largest student body in the United States and a faculty to match, offers Hoag plenty of those. "I can pretty much know that I can get advice or expertise, or even a new collaborator, in just about any new direction my research might move in," she says.

Also important, says Hoag, is "the demeanor of the individuals that I work with. My office, my main colleagues in the diagnostics program faculty—we're a team. Everyone is a team player, very entrepreneurial, always looking for new ways to reach the public. It's so exciting," she adds. "Nobody could possibly become deadwood."

–JIM AUSTIN

Jim Austin is the editor of ScienceCareers.org.